

Die Modellierung zeitvariabler Parameter in diskreten Wahlmodellen

D i s s e r t a t i o n

zur Erlangung des Doktorgrades
der Wirtschaftswissenschaften

vorgelegt von
Dipl.-Kfm. Daniel Guhl
aus Mönchengladbach

genehmigt von der
Fakultät für Energie- und Wirtschaftswissenschaften
der Technischen Universität Clausthal

Tag der mündlichen Prüfung

23. Juli 2014

Vorsitzender der Promotionskommission

Prof. Dr. Wolfgang Pfau

Hauptberichterstatte

Prof. Dr. Winfried Steiner

Berichterstatte

Prof. Dr. Hartwig Steffenhagen

Vorwort

Die vorliegende Dissertationsschrift ist das Ergebnis meiner Forschungsarbeit an der RWTH Aachen und der TU Clausthal. Im Juni 2014 wurde sie von der Fakultät für Energie- und Wirtschaftswissenschaft der TU Clausthal als Dissertation angenommen.

Mein aufrichtiger Dank gilt meinem Doktorvater und akademischen Lehrer, Herrn Prof. Dr. Hartwig Steffenhagen, der mir den Anstoß zum Thema „Dynamik“ gab und meine Arbeit durch seine kritischen und konstruktiven Anregungen und Hinweise gefördert hat. Er ließ mir stets genügend Freiheiten, mich dem Thema auf meine Weise zu nähern. Danken möchte ich ihm auch für die schöne Zeit in Aachen an seinem Lehrstuhl.

Ebenfalls danke ich Herrn Prof. Dr. Winfried Steiner. Er half mir durch zahlreiche Diskussionen und Verbesserungsvorschläge entscheidend in der Endphase der Dissertation und sein Interesse an quantitative Methoden im Marketing war sehr inspirierend. Zudem bin ich ihm für den reibungslosen Ablauf der Promotion an der TU Clausthal zu Dank verpflichtet.

Einen wichtigen Beitrag an meiner Dissertation hatte Prof. Dr. Dieter Urban, der leider bereits im Jahr 2011 verstarb. Es waren seine hervorragenden Doktorandenkurse an der RWTH Aachen, die mein Interesse für ökonometrische Methoden verstärkten und von ihm stammte die Idee, den Kontrollfunktionenansatz zur Behandlung des Endogenitätsproblems in Kapitel 3 zu verwenden.

Danken möchte ich auch allen meinen ehemaligen Kolleg(inn)en in Aachen und Clausthal, die mich in vielfältiger Weise unterstützt haben. Hier sind vor allem Anett Weber, Friederike Paetz und Marcus Kunter zu erwähnen, die mir durch viele Diskussionen zur Methodik und zum Inhalt meiner Dissertation behilflich waren.

Raimund Bau hat nicht nur vor vielen Jahren mein Interesse für diskrete Wahlmodelle geweckt und mich tatkräftig bei den ersten Programmierversuchen zum Kalman Filter unterstützt, er stellte mir auch den Waschmitteldatensatz für die Untersuchung in Kapitel 4 zur Verfügung. Hierfür und für unzählige Diskussionen zur Theorie und Methodik sowie für sämtliche Unterstützung, insb. in der Endphase des Dissertationsprojektes, möchte ich ihm danken. Mein Dank gilt auch Sebastian Gabel, dessen Expertise in der Programmiersprache R sehr hilfreich war, vor allem wenn es darum ging, langsamen R-Code „Beine zu machen“.

Bei meiner Familie bedanke ich mich für all die Jahre der Unterstützung, auch bereits während meines Studiums. Meiner Freundin Juliet, die mit mir gemeinsam alle Höhen und Tiefen dieser Dissertation durchlebt und durchgestanden hat, möchte ich ganz besonders danken. Ohne ihre stetige Motivation, das kritische inhaltliche Feedback und die sorgfältige Durchsicht des Manuskripts, wäre die Arbeit wohl niemals fertig geworden, so dass ich ihr diese Arbeit widmen möchte.

Berlin, im Juli 2014

Daniel Guhl

Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis	vi
1 Einführung	1
1.1 Motivation	1
1.2 Aufbau und Inhalt	5
1.3 Ziel	8
2 Zustandsraummodelle und Kalman Filter im Marketing	11
2.1 Einleitung	11
2.2 Zustandsraummodelle	13
2.3 Zustandsschätzung mit dem Kalman Filter	22
2.4 Parameterschätzung mit der Maximum-Likelihood-Methode	28
2.5 Empirische Studie	32
2.6 Softwareüberblick	41
2.7 Zusammenfassende Schlussbetrachtung	43
Anhang	44
3 Endogenität in Logitmodellen mit zeitvariablen Parametern	49
3.1 Einleitung	49
3.2 Modellierung	53
3.2.1 Nachfragemodell	54
3.2.2 Preismodell	57
3.2.3 Kontrollfunktionenansatz	59
3.2.4 Zustandsraummodell	62
3.3 Schätzung	63
3.3.1 Konstante Parameter	64
3.3.2 Zeitvariable Parameter	68
3.4 Simulationsexperiment	69
3.4.1 Datengenerierender Prozess	70

3.4.2	Ergebnisse der Simulation	72
3.5	Empirische Studie	78
3.5.1	Daten	79
3.5.2	Instrumentalvariablen	81
3.5.3	Ergebnisse	82
3.5.4	Implikationen	89
3.6	Zusammenfassende Schlussbetrachtung	93
4	Die Messung zeitvariabler Markenwerte mit Haushaltspanel-	
	daten	95
4.1	Einleitung	95
4.2	Literaturüberblick	101
4.3	Modellierung	107
4.3.1	Nachfragemodell	107
4.3.2	Endogenitätskorrektur	112
4.4	Schätzung	113
4.4.1	Erste Stufe	114
4.4.2	Zweite Stufe	116
4.5	Empirische Studien	118
4.5.1	Daten	119
4.5.2	Ergebnisse	122
4.5.3	Implikationen	134
4.6	Zusammenfassende Schlussbetrachtung	142
	Anhang	146
	Literaturverzeichnis	148

Abbildungsverzeichnis

2.1	Wirkungsbeziehungen im Zustandsraummodell	15
2.2	Schritte zur rekursiven Berechnung des Kalman Filters	23
2.3	Grafische Darstellung einer Kalman Filter-Iteration	25
2.4	Algorithmus zur Maximum-Likelihood-Schätzung unbekannter (konstanter) Parameter in Zustandsraummodellen	30
2.5	Grafische Darstellung der zeitlichen Parameterverläufe	39
2.6	Grafische Darstellung der zeitlichen Entwicklung der 95 %-Konfidenzintervalle für die Zustandsprognosen der markenspezifischen Grundnutzen	40
3.1	Konzeption der Parameterschätzung	65
3.2	Grafische Darstellung der zeitlichen Preisparameterverläufe ausgewählter Replikationen	77
3.3	Grafische Darstellung der zeitlichen Parameterverläufe	87
4.1	Zeitliche Entwicklung der Markenwerte internationaler Top-Marken	96
4.2	Zeitliche Verteilung der Werte für die fixen Effekte im Margarine- und Waschmittelmarkt	130
4.3	Grafische Darstellung der zeitlichen Markenwertverläufe (Margarine)	131
4.4	Grafische Darstellung der zeitlichen Markenwertverläufe (Waschmittel)	133
4.5	Grafische Darstellung der zeitlichen Zahlungsbereitschaftsverläufe (Margarine)	136
4.6	Grafische Darstellung der zeitlichen Zahlungsbereitschaftsverläufe (Waschmittel)	138
4.7	Nachfragefunktion der Margarinemarke Blue Bonnet	141

Tabellenverzeichnis

1.1 Übersicht zum Inhalt der Arbeit	5
2.1 Trennung kurz- und langfristiger Marketingeffekte	20
2.2 Schätzergebnisse für das MNL- und das TVP-MNL-Modell	36
2.3 Fitmaße für das MNL- und das TVP-MNL-Modell	38
3.1 Modellübersicht und Forschungsfragen	71
3.2 Resultate des Monte Carlo-Experiments (konst. Parameter)	73
3.3 Resultate des Monte Carlo-Experiments (zeitvar. Parameter)	76
3.4 Deskriptive Statistiken zu den Thunfischdaten	80
3.5 MSL-Schätzergebnisse für die untersuchten Modelle	85
3.6 Modellvergleich auf Basis der Preisoptimierung	91
4.1 Systematisierung alternativer Maße des Markenwerts	97
4.2 Ausgewählte Beiträge zur Abbildung von Dynamik in diskreten Wahlmodellen für Individualdaten	102
4.3 Ausgewählte Beiträge zur Verwendung fixer Effekte in diskreten Wahlmodellen für Individualdaten	105
4.4 Deskriptive Statistiken zu den markenspezifischen Variablen der Margarine- und Waschmitteldaten	119
4.5 Deskriptive Statistiken zu den Haushalten und deren Einkaufsverhalten in der Margarine- und Waschmittelkategorie	120
4.6 Modellvergleich auf der ersten Stufe	123
4.7 Schätzergebnisse für die konst. Parameter (Margarine)	126
4.8 Schätzergebnisse für die konst. Parameter (Waschmittel)	128
4.9 Verteilung der Wechselkosten	139
4.10 Schätzergebnisse für die Kovarianzmatrix (Margarine)	146
4.11 Schätzergebnisse für die Kovarianzmatrix (Waschmittel)	146
4.12 Schätzergebnisse für die Preisregression (Margarine)	147
4.13 Schätzergebnisse für die Preisregression (Waschmittel)	147

Kapitel 1

Einführung

„Understanding and measuring the effects of consumer choice is one of the richest and most challenging aspects of research in marketing.“

Chandukala et al. (2007, S. 99)

„Markets are dynamic by nature, and marketing efforts can be directed to stimulate, reduce, or to utilize these dynamics.“

Leeflang et al. (2009, S. 13)

1.1 Motivation

Die *Analyse der Konsumentennachfrage* spielt vor allem im *Marketing* seit jeher eine *bedeutende Rolle* (Chintagunta und Nair 2011). Um Marketinginstrumente optimal einsetzen zu können, ist es aus *Unternehmenssicht* wichtig zu verstehen, warum Konsumenten, zu welchem Zeitpunkt, welche Entscheidungen treffen. Diese Aufgabe ist äußerst *komplex* und bis heute *Gegenstand der Marketingforschung* (Allenby 2012).

Bei der Nachfrageanalyse gilt es, eine Reihe wichtiger Konsumenteneigenschaften zu berücksichtigen, welche im Marketing und angrenzenden Fachrichtungen erforscht und belegt werden konnten. Zudem haben verschiedene Datenquellen gewisse Eigenheiten, die es zu beachten gilt.

1. Konsumenten treffen Entscheidungen nicht zufällig, sie handeln *zielgerichtet* (Chintagunta und Nair 2011). Unter Einsatz knapper Ressourcen versuchen sie, bestimmte Vorhaben umzusetzen und sich

hierdurch besser zu stellen. Solche Entscheidungsregeln sollten in die Modellierung einfließen, um theoriegestützte und damit robuste Aussagen treffen zu können (Reiss 2011).

2. Nicht alle Konsumenten sind gleich (Allenby und Rossi 1999). Tatsächlich unterscheiden sich sämtliche Konsumenten in ihren Präferenzen und Bedürfnissen voneinander. Äußere und situative Einflüsse bestimmen ebenfalls das individuelle Handeln. Eine Vernachlässigung dieser *Heterogenität* der Konsumenten führt zu falschen Ergebnissen (Jain *et al.* 1994).
3. Marktdaten stellen eine hervorragende Informationsbasis zur Modellierung dar. Im Gegensatz zu Befragungsdaten sind sie 1.) heutzutage in aggregierter und disaggregierter Form reichhaltig und direkt verfügbar (Leeflang 2011), 2.) nicht betroffen von Verzerrungen aufgrund fehlender Anreizkompatibilität (Miller *et al.* 2011) und 3.) reliabel (Guadagni und Little 1983). Marktdaten weisen aber ein mögliches *Endogenitäts*problem auf, da nicht nur Nachfrager, sondern auch Anbieter zielgerichtet handeln (Allenby 2012) und der Analytiker nicht alle Variablen beobachten kann (Petrin und Train 2010). Folglich ist die explizite Behandlung dieses Problems bei Nachfrageanalysen ratsam.
4. *Dynamik* ist im Marketing entscheidend, denn Anbieter und Nachfrager treffen auf Märkten selten in einem statischen Gefüge zusammen (Hanssens und Parsons 1993). Konsumentenpräferenzen ändern sich über die Zeit, z. B. beeinflusst durch Neuprodukteinführungen, saisonale Effekte, wechselnde Trends oder neue rechtliche Rahmenbedingungen (Lachaab *et al.* 2006). Marketingaktivitäten besitzen darüber hinaus langfristige und sich im Zeitablauf ändernde Wirkungen (Leeflang *et al.* 2009).

In der vorliegenden Arbeit werden *diskrete Wahlmodelle* behandelt.¹ Hierbei handelt es sich um vielfältig einsetzbare *mathematische Nachfragemodelle*, welche *deskriptive*, *prädiktive* sowie *normative* Analysen erlauben und im Marketing wegen ihrer *inhaltlichen* und *methodischen Stärken* äußerst beliebt sind (Chandukala *et al.* 2007; Chintagunta und Nair 2011).

¹ In den letzten Jahrzehnten wurden viele verschiedene Modelle zur Nachfrageanalyse vorgeschlagen und einen umfassenden Überblick hierzu liefern z. B. die Lehrbücher von Leeflang *et al.* (2009) und Hanssens *et al.* (2001). Neuere Entwicklungen fassen bspw. Leeflang und Hunneman (2010) zusammen.

Diese Modelle bilden das Nachfrageverhalten einzelner Konsumenten bei ihren *diskreten Auswahlentscheidungen* zwischen verschiedenen zur Verfügung stehenden Optionen ab. Dabei wird ein *ökonomisch fundiertes Entscheidungskalkül* unterstellt: Konsumenten wählen Alternativen, die ihnen den höchsten Nutzen stiften (McFadden 2001). Dies ist für die Nachfrageanalyse aus mehreren Gründen vorteilhaft: Zum einen sind Entscheidungen von Konsumenten (insb. im Bereich schnellldrehender Konsumgüter) meist Wahlentscheidungen zwischen diskreten Optionen – z. B. die Wahl eines bestimmten Joghurts aus dem Supermarktsortiment. Zum anderen lässt sich dieses Modell auf Konsumentenebene sowohl auf *Individualdaten* (Wahlentscheidungen einzelner Konsumenten) als auch auf *aggregierte Daten* (Marktanteile) anwenden (Dubé *et al.* 2002). Folglich ist eine Trennung der Modellierungs- und Datenebene möglich und zudem gibt es inzwischen Erweiterungen bei der Modellschätzung zur Behandlung von *Endogenität* (Villas-Boas und Winer 1999). Weiterhin erlaubt die Modellierung auf Individualebene die Abbildung (unbeobachteter) *Konsumentenheterogenität* (Allenby und Rossi 1999). Und schließlich liefert die Annahme der Nutzenmaximierung bessere Prognosen für bisher nicht in den Daten abgebildete Nachfrageszenarien, weil das Zustandekommen der Konsumentenentscheidung explizit modelliert wird (Chintagunta und Nair 2011).

Die oben angeführten Punkte eins bis drei werden von gängigen diskreten Wahlmodellen adäquat berücksichtigt (Chintagunta *et al.* 2005). Dem Thema *Dynamik* wird allerdings bislang unzureichend Beachtung geschenkt,² obwohl ein zunehmender Trend erkennbar ist (Pauwels *et al.* 2004; Leeflang *et al.* 2009; Dekimpe und Hanssens 2010). Die vorliegende Arbeit setzt genau am *vierten Punkt* an. Sie behandelt *diskrete Wahlmodelle mit zeitvariablen Parametern* und zeigt deren Nützlichkeit für die Marketingforschung und -praxis.³

1. In vielen Situationen verlangt der *Untersuchungsgegenstand* zeitlich flexible Modellstrukturen, da sich andernfalls bestimmte Fragestellungen nicht beantworten lassen (Leeflang *et al.* 2009). Ferner besteht ein generelles Interesse daran, zu verstehen, wie sich Konsumentenpräferenzen über die Zeit verändern (Lachaab *et al.* 2006).

² Ausnahmen stellen bspw. Lachaab *et al.* (2006) oder Rutz und Sonnier (2011) dar.

³ Die in der vorliegenden Arbeit behandelten diskreten Wahlmodelle, welche Dynamik in Form von zeitvariablen Nutzenparametern berücksichtigen, sind von *dynamischen diskreten Wahlmodellen* zu unterscheiden. Letztere bilden ein *vorausschauendes Entscheidungsverhalten* der Konsumenten ab (vgl. hierzu z. B. Baumgartner 2000, Erdem *et al.* 2003 oder Hendel und Nevo 2006).

2. Das Ignorieren zeitvariabler Effekte in Nachfragemodellen kann zu Verzerrungen in Modellkomponenten führen, die dann diese Dynamik irrtümlich aufnehmen (Harvey 1989). Hieraus würden sich fehlleitende Implikationen für das Management ergeben.
3. Modelle mit zeitvariablen Parametern liefern i. d. R. einen besseren *Datenfit* (Putsis 1998). Entsprechend präzisere Ergebnisse führen zu besseren Implikationen für das Management, da die modellbasierte Entscheidungsunterstützung weniger unsicherheitsbehaftet ausfällt.
4. Vor dem Hintergrund üblicher *Längsschnittdimensionen* der heute verfügbaren Daten von mehreren Jahren (Pauwels *et al.* 2004), erscheint eine gewisse Skepsis gegenüber zeitlich „stabiler“ Modelle berechtigt. Anstatt dann nur den zeitlich jüngeren Teil der Daten zu verwenden und vorhandene (ältere) Informationen unberücksichtigt zu lassen, erlauben Modelle mit zeitvariablen Parametern eine angemessenere Lösung dieses „Datenproblems“.
5. Vor allem *Praktiker* mögen den „Blick in den Rückspiegel“ nicht sonderlich gerne (Simon und Fassnacht 2009). Adaptive Modelle haben eine höhere Vertrauenswürdigkeit, weil diese aktuell (also am Ende der Daten) das „bestmögliche“ Ergebnis liefern und nicht nur eines, das zur (durchschnittlichen) Vergangenheit passt.
6. Die sog. *Lucas-Kritik*, welche besagt, dass Konsumenten ihr Verhalten an die „Rahmenbedingungen“ anpassen, lässt sich durch die Verwendung zeitvariabler Parameter „abmildern“, weil das Kernproblem explizit thematisiert wird (van Heerde *et al.* 2005), ohne restriktive Annahmen zu benötigen (Amman und Kendrick 2003).

Weiterhin kommen in der vorliegenden Arbeit *stochastisch* zeitvariierende Parameter zu Einsatz, weshalb beim Analytiker weniger Vorwissen über Dynamiken vorhanden sein muss. Hierzu formuliert Putsis (1998, S. 233) treffend: „*What should one do in the case of weak prior information on the form of the variation over time? Stochastic parameter models are ideally applied in such situations since the temporal pattern of parameter variation is determined from the data, as opposed to being imposed on the data.*“ Hierdurch lassen sich die Modelle breiter einsetzen und allgemeinere Erkenntnisse gewinnen. Solche diskreten Wahlmodelle sind bis heute sehr selten und dementsprechend wurden diese auch noch nicht hinreichend untersucht.⁴

⁴ Bestehenden Forschungsbedarf äußern bspw. auch die Editoren der aktuellen *Special Section on „Marketing Dynamics“* des *Journal of Marketing Research* (JMR). Dort betonen

1.2 Aufbau und Inhalt

Die vorliegende Arbeit besteht im Kern aus *drei Kapiteln* (2, 3 und 4), wobei jedes Kapitel einen *eigenständigen Beitrag* darstellt und auch als solcher zu verstehen ist. Da der erste Beitrag eine *methodische Einführung* in die *grundsätzliche Vorgehensweise* der nachfolgenden Kapitel liefert, wird die Lektüre des 2. Kapitels *vor* den Kapiteln 3 oder 4 empfohlen. Die beiden Beiträge der Kapitel 3 und 4 sind prinzipiell unabhängig voneinander. Sie vertiefen methodisch das Vorgehen aus dem ersten Beitrag und wenden komplexere Modellvarianten an. Tabelle 1.1 liefert eine synoptische Übersicht zum Inhalt der Arbeit. Eine detailliertere Beschreibung der einzelnen Beiträge ist den darauf folgenden Ausführungen zu entnehmen.

Tabelle 1.1: Übersicht zum Inhalt der Arbeit

	Kapitel 2 (Beitrag 1)	Kapitel 3 (Beitrag 2)	Kapitel 4 (Beitrag 3)
Gegenstand	Dynamische Marketingmodelle	Preisendogenität in dyn. Modellen	Dynamische Messung von Markenwerten
Daten	aggregiert	aggregiert	disaggregiert
untersuchte Produktkategorie	Thunfisch	Thunfisch	Margarine und Waschmittel
Heterogenität	nein	Preisparameter	alle Parameter
Preisendogenität	nein	ja	ja
Dynamik	alle Parameter	alle Parameter	Markenparameter
Simulationsstudie	nein	ja	nein
Normative Analyse	nein	ja	nein

Beitrag 1: Zustandsraummodelle und Kalman Filter im Marketing

Der erste Beitrag (Kapitel 2) führt zunächst in sog. *Zustandsraummodelle* und den *Kalman Filter* ein (Petrís *et al.* 2009). Diese modernen Werkzeuge zur Modellierung und Schätzung dynamischer Zusammenhänge besitzen viele vorteilhafte Merkmale, die sie für den Einsatz im Marketing prädestinieren: Sie ermöglichen z.B. die Abbildung und Schätzung potenziell nichtstationärer dynamischer Zusammenhänge, die Berücksichtigung von Messfehlern in den Variablen und die Behandlung fehlender Werte.

Bronnenberg *et al.* (2012, S. 749): „We [...] invited papers that study the evolution of consumer preferences and the dynamic effects of marketing variables [...]. We were motivated by the fact that JMR receives relatively few submissions in this important research area [...].“

Um die Beziehung zu vielen marketingrelevanten Problemen und Fragestellungen aufzuzeigen, wird im ersten Beitrag für mehrere dynamische Marketingmodelle die Zustandsraumform präsentiert und Einzelheiten erörtert. Es wird detailliert erklärt, wie der rekursive Kalman Filter-Algorithmus zur *Schätzung der dynamischen Zustände* arbeitet und wie hierzu die vorhandene Information der Daten von Periode zu Periode verwendet wird. Der Kalman Filter ist weiterhin hilfreich zur Aufstellung der Likelihoodfunktion des Modells und bildet damit auch das Fundament zur *Schätzung konstanter Modellparameter*.

Eine *empirische Studie* mit Bezug zur dynamischen nutzenbasierten Marktanteilsmodellierung (inkl. frei zugänglicher Daten und R-Code (R Core Team 2013)) illustriert, welche Schritte bei der Anwendung der Methode notwendig sind und welche Vorzüge eine dynamische Modellierung besitzt. Neue Ansätze sind nur dann praktikabel, sofern es Computerprogramme gibt, die den problemlosen Einsatz ermöglichen. Nicht alle Programme besitzen allerdings denselben Funktionsumfang und lassen sich gleich gut bedienen. Daher beleuchtet ein kritischer *Softwareüberblick*, welche Programme sich insb. für den anwendungsorientierten Einsatz im behandelten Marketingkontext eignen. Aktuelle Erweiterungen zu Zustandsraummodellen und dem Kalman Filter werden in der *Schlussbetrachtung* aufgezeigt.

Beitrag 2: Endogenität in Logitmodellen mit zeitvariablen Parametern

Der zweite Beitrag (Kapitel 3) behandelt *Endogenität in Logitmodellen mit zeitvariablen Parametern für aggregierte Daten*. Korreliert die Preisvariable mit dem Störterm des Modells, wird von Preisendogenität gesprochen und dies stellt bekanntlich in Nachfragemodellen, die auf nichtexperimentellen Daten basieren, ein typisches Problem dar. Eine Vernachlässigung oder falsche Handhabung dieses Problems führt zu verzerrten Ergebnissen, unabhängig vom Umfang der verwendeten Datenbasis. Ist der Preisparameter zeitvariabel, versagen etablierte Methoden zur Endogenitätskorrektur.

Es wird ein *Kontrollfunktionenansatz* vorgestellt, ein Verfahren, das in der beschriebenen Situation trotzdem einsetzbar ist (Kim und Kim 2011). Die Verwendung des Kontrollfunktionenansatzes in Verbindung mit Zustandsraummodellen und dem Kalman Filter erlaubt die erwartungstreue Schätzung des Modells, was durch eine umfangreiche *Simulationsstudie* belegt wird. Es zeigt sich allerdings, dass das Ignorieren der Endogenität in manchen Fällen sogar zu besseren Ergebnissen führt, als eine falsche Berück-

sichtigung. Die Behandlung der Endogenität mit dem Kontrollfunktionsansatz ist jedoch stets die beste Entscheidung.

Eine *empirische Studie* verdeutlicht die Anwendung des Kontrollfunktionsansatzes mit aggregierten Daten auf Händlerebene. Es gibt klare zeitliche Schwankungen in allen Nutzenparametern und die Preissensitivitäten sind endogenitätsbedingt verzerrt. Neben der statistischen Signifikanz der Endogenität wird abschließend mithilfe einer *normativen Analyse* ebenfalls ihre betriebswirtschaftliche Bedeutsamkeit untersucht und belegt. Der betrachtete Händler erwirtschaftet unter Verwendung des Kontrollfunktionsansatzes einen mehr als drei Prozent höheren Deckungsbeitrag im Vergleich zu einer Situation mit einem Modell ohne Endogenitätsberücksichtigung. Dieser Wert ist durchaus beachtlich, bei jährlichen Deckungsbeiträgen in der betrachteten Kategorie (Dosenthunfisch) von mehr als einer halben Mio. \$. Ein Ausblick mit möglichen Erweiterungen für zukünftige Forschungsarbeiten ist in der *Schlussbetrachtung* zu finden.

Beitrag 3: Die Messung zeitvariabler Markenwerte mit Haushaltspanel-daten

Das 4. Kapitel behandelt die *dynamische nutzenbasierte Messung des Markenwerts* mit *Haushaltspanel-daten*. Marketingpraktiker und -forscher gleichermaßen nutzen das Konzept des Markenwerts, um auszudrücken, dass *Marken* zu einer Produktdifferenzierung führen. Selbst technisch/funktional vergleichbare Produkte werden von Konsumenten unterschiedlich wahrgenommen, sofern sie kognitiv mit einer Marke(npersönlichkeit) in Verbindung gebracht werden, die implizit für das Produkt bürgt. Damit besitzen Marken eine zentrale Funktion im Marketing, und Hersteller investieren hohe Summen in deren Aufbau, Pflege und Schutz. Selbst wenn die Relevanz von Marken außer Frage steht, ist bis heute unklar, wie genau der Markenwert gemessen werden soll, und in der Literatur existieren viele verschiedene Definitionen und Metriken (Farris *et al.* 2010).

Kamakura und Russell (1993) schlugen bereits vor 20 Jahren die Markenwertmessung mithilfe *diskreter Wahlmodelle* vor. Der Ansatz besitzt viele günstige Eigenschaften, z. B. die ökonomische Fundierung oder die Berücksichtigung von Konsumentenheterogenität, vernachlässigt allerdings bislang die mögliche *Zeitvariabilität* von Markenwerten.

Im dritten Beitrag wird ein *zweistufiges Vorgehen* präsentiert, das die flexible Messung zeitvariabler Markenwerte erlaubt, und zwar unter Berücksichtigung von (Preis-)Endogenität, Kauffeedbackeffekten sowie unbeob-

achteter Konsumentenheterogenität und Markencharakteristika. Auf der ersten Schätzstufe erfolgt zunächst eine Trennung der Effekte auf Individualebene (z. B. Heterogenität und Kauffeedbackeffekte) und aggregierter Ebene (z. B. Markenwertdynamik und Preisendogenität). Hierzu wird ein Ansatz mit *fixen Effekten* verwendet (Chintagunta *et al.* 2005). Auf der zweiten Stufe lässt sich ein Zustandsraummodell formulieren und die zeitvariablen Markenwerte werden über den Kalman Filter bzw. Kalman Smoother ermittelt (Harvey 1989).

Zwei empirische Studien illustrieren die Leistungsfähigkeit des postulierten Ansatzes für Produktkategorien aus dem Bereich schnelldrehender Konsumgüter (Margarine und Waschmittel). Es zeigt sich, dass Markenwerte in der Tat über die Zeit variieren und die zeitlichen Unterschiede monetär ausgedrückt bis zu 30 % des durchschnittlichen Kategoriepreises ausmachen. Vor allem praxisorientierten Anwendern wird damit ein Werkzeug bereitgestellt, mit dem sie schnell und einfach vorhandene Haushaltspaneldaten besser nutzen können, um interessante und vor allem relevante Einsichten bzgl. dynamischer Entwicklungen der Marken zu gewinnen.

1.3 Ziel

Die vorliegende Arbeit verfolgt eine Reihe von *Zielen*:

- Sie soll zum Verständnis beitragen, wie *diskrete Wahlmodelle* mit expliziter Berücksichtigung von Konsumentenheterogenität, Preisendogenität sowie zeitvariablen Parametern arbeiten und welche Stärken und Schwächen sie besitzen. Erst hierdurch wird eine objektive Abwägung möglich, welcher Komplexitätsgrad in der Modellierung für eine gegebene Fragestellung angebracht ist. Das passt zur Forderung von Chintagunta *et al.* (2013, S. 7): „*We must learn to recognize when hot-button issues [endogeneity, heterogeneity, time-varying parameters . . .] are relevant and when they are not.*“
- Dem Leser soll der Kalman Filter als flexible *Methode* zur Schätzung dynamischer Zusammenhänge verständlich vermittelt werden. Besonders in Verbindung mit Marketingproblemen gilt es, Anwendungsmöglichkeiten zu verstehen, die von der üblichen Zerlegung univariater Zeitreihen abweichen. Da die Anzahl der Publikationen (insb. im deutschsprachigen Raum) noch spärlich ist, soll die Arbeit den Gebrauch und die Verbreitung des Kalman Filters fördern.

- Die empirischen Studien sind einerseits als „proof of concept“ zu verstehen, andererseits gilt es hiermit zudem, Konsumenten sowie die Art und Weise, wie sich deren Präferenzen über die Zeit ändern, besser zu verstehen. Es geht also auch um die Gewinnung *inhaltlicher* Einsichten, die besonders die betriebswirtschaftlichen Konsequenzen von z. B. zeitvariierenden Preissensitivitäten oder Markenwerten betreffen.

Es besteht folglich das Anliegen, einen *modelltechnischen, methodischen* und *inhaltlichen* Beitrag zur Dynamik in der Analyse der Konsumentennachfrage zu leisten.

Kapitel 2

Zustandsraummodelle und Kalman Filter im Marketing

„Marketing for a firm rarely takes place in a static environment.“

Hanssens und Parsons (1993, S. 422)

2.1 Einleitung

Märkte sind *dynamisch*. Produktinnovationen, wechselnde Konkurrenzsituationen, Trends und saisonale Effekte führen dazu, dass sich Konsumentenpräferenzen fortwährend ändern (Lachaab *et al.* 2006). Viele Marketingaktivitäten besitzen zudem langfristige und zeitvariierende Wirkungen. Für Unternehmen ist es daher wichtig, diese Marketingdynamiken zu kennen und zu nutzen, um nachhaltige Wettbewerbsvorteile zu erzielen (Leeflang *et al.* 2009).

Infolgedessen sind *dynamische Modelle* zur Untersuchung marketingrelevanter Fragestellungen sehr gebräuchlich und besitzen eine lange Tradition (Dekimpe und Hanssens 2000, S. 183 ff.). Die *empirische Schätzung* solcher Modelle wird heutzutage begünstigt durch reichhaltige Daten (gering aggregiert mit vielen Beobachtungen in Längs- und Querschnittsdimension) sowie moderne zeitreihenanalytische und ökonometrische Ansätze (Dekimpe und Hanssens 2000; Pauwels *et al.* 2004). Es verwundert daher nicht, dass jüngst die Anzahl an Publikationen zur „Marketing-Dynamik“ stetig zugenommen hat (Leeflang *et al.* 2009).

Gegenstand des vorliegenden Beitrags ist der *Kalman Filter* (kurz: KF) und die Darstellung dessen Leistungsfähigkeit für eine datengestützte dynamische Analyse von Marketingphänomenen. Der KF ist eine statistische Methode zur Schätzung dynamischer Modelle, die auf *Zustandsraummodellen* (kurz: ZRM) basiert. Dieser Methode fehlt im Marketing noch eine weite Verbreitung, obwohl sie sich aufgrund ihrer expliziten Berücksichtigung dynamischer Zusammenhänge für die Behandlung oben genannter Sachverhalte besonders gut eignet. ZRM zusammen mit dem KF besitzen eine Reihe vorteilhafter Eigenschaften, die sie insb. für Anwendungen im Marketing prädestinieren, da sie dort häufig anzutreffende Herausforderungen (z. B. Messungenauigkeiten, nichtstationäre Daten, fehlende Werte, latente Variablen, ...) in einem einheitlichen Modellierungs- bzw. Schätzansatz handhaben (van Heerde *et al.* 2004, S. 167 f.). Diese Leistungsfähigkeit des Ansatzes ist der Hauptgrund, warum er sich neuerdings in der internationalen Marketingforschung steigender Beliebtheit erfreut, wie die Überblicksartikel von Pauwels *et al.* (2004), Dekimpe *et al.* (2008) und Leeftang *et al.* (2009) sowie die dort zitierten Arbeiten belegen.

Der vorliegende Beitrag verfolgt drei *Ziele*: Er soll

1. interessierten Lesern (insb. Studierenden und Doktoranden sowie dem Marketing-Forschungspersonal in Unternehmen) Grundlegendes zu ZRM und zum KF verständlich erläutern,
2. durch Beispiele mit Marketingbezug den Zugang hierzu erleichtern sowie
3. zusätzlich die Möglichkeiten und die Leistungsfähigkeit des Ansatzes eingängig vermitteln.

Folgender *Aufbau* wird gewählt: Da die Formulierung eines Modells als ZRM notwendig ist, um den KF zur Schätzung dynamischer Modellkomponenten (z. B. zeitvariable Parameter) nutzen zu können, werden zunächst (Abschnitt 2.2) ZRM ausführlich an marketingrelevanten Beispielen vorgestellt. Der KF wird in Abschnitt 2.3 präsentiert und ebenfalls an einem einfachen Beispiel erläutert. Abschnitt 2.4 thematisiert die Maximum-Likelihood-Schätzung unbekannter Parameter, die nicht über den KF ermittelt werden. Hiernach (Abschnitt 2.5) folgt eine empirische Studie mit Marketingbezug und Abschnitt 2.6 diskutiert unterschiedliche Softwarepakete zur Anwendung des KF. Der Beitrag schließt in Abschnitt 2.7 mit einer Zusammenfassung sowie einem Ausblick.

2.2 Zustandsraummodelle

ZRM haben ihren Ursprung in den Ingenieurwissenschaften und wurden u. a. zur Positionsbestimmung von Flugobjekten entwickelt (Kalman 1960). Sie sind äußerst flexibel einsetzbare Modelle, die sehr viele Modelle bzw. Modellklassen nesten können (Shumway und Stoffer 2011). Wie bereits erwähnt, ist die Zustandsraumform eines dynamischen Modells die Grundlage für die Anwendung des KF.

Im allgemeinen *linearen Fall*¹ lautet das ZRM in Matrixform (Harvey 1989, S. 100 ff.):²

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{a}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{R}_t) \text{ und} \quad (2.1)$$

$(m \times 1) \quad (m \times n)(n \times 1) \quad (m \times 1) \quad (m \times 1) \quad (m \times 1)(m \times m)$

$$\mathbf{a}_t = \mathbf{F}_t \mathbf{a}_{t-1} + \mathbf{c}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \quad \forall t. \quad (2.2)$$

$(n \times 1) \quad (n \times n)(n \times 1) \quad (n \times 1) \quad (n \times 1) \quad (n \times 1)(n \times n)$

Gleichung (2.1) ist die *Beobachtungsgleichung* und der Beobachtungsvektor $\mathbf{y}_t = [y_{1t} \ y_{2t} \ \dots \ y_{mt}]'$ enthält die zu jedem Zeitpunkt t *tatsächlich beobachteten* m Werte der gemessenen dynamischen Größe. Denkbar sind z. B. Absätze oder Bekanntheitswerte mehrerer Marken. Gleichung (2.2) ist die *Zustandsgleichung*. Der Zustandsvektor $\mathbf{a}_t = [a_{1t} \ a_{2t} \ \dots \ a_{nt}]'$ beinhaltet die *wahren, aber unbeobachtbaren* n Werte der gemessenen dynamischen Größe. Die genaue Interpretation dieser Zustände ist von der jeweiligen Anwendung abhängig. Denkbar sind u. a. dynamische latente Konstrukte wie Goodwill oder auch zeitvariable Parameter. Näheres hierzu wird in den nachfolgenden Beispielen erörtert. Die Zustandsgleichung wird als eine zeitdiskrete *Differenzengleichung 1. Ordnung* spezifiziert. Hierdurch wird Dynamik mithilfe der *Übergangsmatrix* \mathbf{F}_t modelliert, und \mathbf{a}_t hängt ausdrücklich von \mathbf{a}_{t-1} , dem Zustand der Vorperiode, ab. \mathbf{H}_t , die *Beobachtungsmatrix*, verbindet die Gleichungen (2.1) und (2.2) miteinander. Die beiden *Driftvektoren* \mathbf{c}_t und \mathbf{d}_t bilden Effekte exogener Größen ab, die additiv auf \mathbf{y}_t bzw. \mathbf{a}_t wirken, z. B. Marketingaktivitäten oder saisonale Effekte. $\boldsymbol{\varepsilon}_t$ und $\boldsymbol{\eta}_t$ sind zwei *Störgrößen*, die den Einfluss nicht berücksichtigter Effekte und Messungenauigkeiten explizit im Modell darstellen. Sie sind

¹ Lineare ZRM mit normalverteilten Fehlern werden in Quellen, die die *bayes'sche Schätztheorie* verwenden, oftmals als *Dynamische Lineare Modelle* (DLM) bezeichnet. Petris et al. (2009, S. 31 ff.) liefern hierzu und zur Historie von ZRM eine Einführung.

² Zur Notation: Fettgedruckt werden Vektoren (kleine Buchstaben) und Matrizen (große Buchstaben) dargestellt. Bei einer kursiven Schreibweise handelt es sich um geschätzte Größen.

normalverteilt mit den jeweiligen Kovarianzmatrizen \mathbf{R}_t bzw. \mathbf{Q}_t und besitzen einen Mittelwert von Null. Ferner sind sie *intertemporal unabhängig* (keine Autokorrelation) und *miteinander* sowie *mit dem Anfangszustand* \mathbf{a}_0 *unkorreliert*. Für den Vektor \mathbf{a}_0 , mit dem das dynamische System initialisiert wird, gilt $\mathbf{a}_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$.

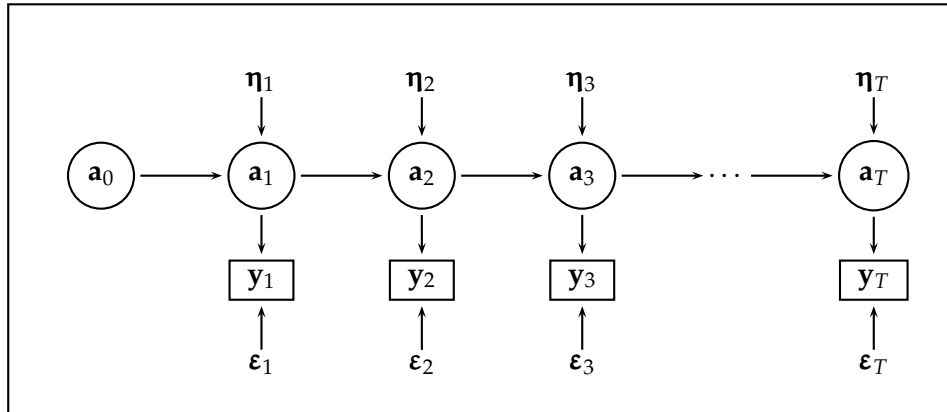
Die *Linearität* des ZRM bezieht sich darauf, dass die Gleichungen (2.1) und (2.2) linear im Zustandsvektor \mathbf{a} sind. Aus zwei Gründen hat diese scheinbar restriktive Annahme keine weitreichenden Auswirkungen.

1. Die Annahme ist nicht besonders streng. Multiplikative dynamische Modelle lassen sich durch Logarithmieren *exakt* linearisieren. Andere Formen nichtlinearer Verknüpfungen in \mathbf{a} (z. B. bei Diffusionsmodellen, vgl. ausführlich Xie *et al.* 1997) können stets mithilfe entsprechender *Approximationen* linearisiert werden. Naik *et al.* (2008) nutzen z. B. den *Extended Kalman Filter* (EKF), der auf einem mit einer Taylor-Reihe 1. Ordnung linearisierten ZRM basiert.
2. Effekte exogener Größen können in den Gleichungen (2.1) und (2.2) problemlos nichtlinear im Zusammenhang und sogar in den Parametern spezifiziert werden. Dies ist für die Verwendung von ZRM im Marketing entscheidend, da *Marketing-Mix-Effekte* generell *nichtlinearer Natur* sind (Leeflang *et al.* 2000, S. 79), so dass hierdurch sinnvolle *normative Aussagen* erst ermöglicht werden (Albers 2012, S. 112).

Alle Vektoren und Matrizen des ZRM besitzen den Zeitindex t . Sie müssen für die Anwendbarkeit des im nachfolgenden Abschnitt beschriebenen KF demnach nicht zeitlich invariant sein. Zeitvariabilität bezieht sich in diesem Modell daher auf sämtliche Größen. Für alle Modellkomponenten außer \mathbf{a}_t (und damit \mathbf{y}_t) muss die zeitliche Variation allerdings *prädeterminiert* sein, weil im ZRM nur die Dynamik von \mathbf{a}_t modellendogen ist (Harvey 1989, S. 101).

Die Beobachtungsgleichung (2.1) trägt der Tatsache Rechnung, dass viele Größen – insb. im Marketing – nur ungenau beobachtbar sind (Naik *et al.* 2007). Messungenauigkeiten (ϵ_t) sind dabei intertemporal voneinander unabhängig, d. h. Messstörungen in t beeinflussen nicht die Zukunft der dynamischen Größe \mathbf{a}_t . Störungen, die diese dynamische Größe selbst betreffen (η_t), sind zwar ebenfalls intertemporal unkorreliert, sie werden aber über die dynamische Zustandsgleichung (2.2) von Periode zu Periode zeitlich übertragen und besitzen folglich einen *nachhaltigen Effekt*.

Abbildung 2.1: Wirkungsbeziehungen im Zustandsraummodell



Quelle: Eigene Darstellung in Anlehnung an Petris et al. (2009, S. 41)

Abbildung 2.1 stellt die Wirkungsbeziehungen im ZRM grafisch dar³ und verdeutlicht zwei wichtige Eigenschaften (Petris et al. 2009, S. 39 ff.):

- *Bedingte Unabhängigkeit:* Die Beobachtungsgleichung kann als ein *reflektives Messmodell* für das *dynamische latente Konstrukt a* interpretiert werden. Folglich sind über *a* die Beobachtungen periodenübergreifend miteinander verbunden. Es ist aber auch ersichtlich, dass die Beobachtungen *gegeben a unabhängig* voneinander sind.
- *Markov-Eigenschaft:* Die gesamte Dynamik des Modells wird über die Zustände abgebildet. Zur Erstellung einer Prognose zum Zeitpunkt *t* ist es allerdings nicht notwendig, alle Zustände (*a*₀, ..., *a*_{t-1}, *a*_t) zu kennen. *a*_t beinhaltet die *gesamte Information der Vergangenheit*.

Aus Abbildung 2.1 wird auch klar, warum es mitunter problematisch sein kann, dynamische Modelle mit „herkömmlichen“ Methoden (z. B. OLS) zu schätzen. Im klassischen linearen Regressionsmodell ist aufgrund der angenommenen Unabhängigkeit der Beobachtungen eine Trennung der Störterme nicht notwendig. Messungenauigkeiten können im „normalen“ Störterm aufgefangen werden (Greene 2011, S. 53). In dynamischen Modellen führt eine fehlende Berücksichtigung von Messfehlern (ϵ_t) dagegen zu *verzerrten Parametern* (Naik et al. 2007, S. 258). Die Formulierung des ZRM ist also nicht nur intuitiv sinnvoll, sondern sie ist essentiell für die Qualität der Ergebnisse.

³ Aus Gründen der Übersichtlichkeit unterbleibt die Darstellung der Driftvektoren.

Zur Veranschaulichung des vorgestellten ZRM zeigen nun zwei Beispiele, wie typische dynamische Marketingmodelle in die passende Zustandsraumform zu überführen sind.

Beispiel 1: Dynamisches Werbewirkungsmodell

Das Modell von Nerlove und Arrow (1962) ist ein sehr beliebtes Modell zur Abbildung dynamischer Werbeeffekte. Es postuliert, dass das Werbebudget x_t einen Goodwill-Stock g_t (ähnlich einem Kapitalstock) aufbaut, der nicht schlagartig, sondern im Zeitablauf nach und nach verschwindet, wenn nicht mehr in Werbung „investiert“ wird. Mit einer linearen Messgleichung für das dynamische Konstrukt Werbegoodwill folgt:⁴

$$y_t = g_t + \varepsilon_t \text{ und} \quad (2.3)$$

$$g_t = \gamma g_{t-1} + \beta x_t + \eta_t, \quad \forall t. \quad (2.4)$$

Der Carry-Over-Koeffizient γ , mit $0 < \gamma < 1$, gibt an, wie viel Prozent vom Goodwill-Stock von Periode zu Periode erhalten bleibt. β ist die marginale Wirkung des eingesetzten Werbebudgets auf den Goodwill-Stock. Die Zustandsraumform ist leicht ersichtlich: mit $\mathbf{H} = 1$, $\mathbf{a}_t = g_t$, $\mathbf{d} = 0$, $\mathbf{R} = \sigma_\varepsilon^2$, $\mathbf{F} = \gamma$, $\mathbf{c}_t = \beta x_t$ und $\mathbf{Q} = \sigma_\eta^2$ wird das Modell gemäß der Gleichungen (2.1) und (2.2) aufgestellt. Betont sei, dass empirische Untersuchungen mit anderen Ansätzen zur Modellierung von Stock-Modellen aus methodischen Gründen oft *eine* der hier in *beiden* Gleichungen des Modells vorkommenden Störgrößen vernachlässigen (vgl. z. B. Köhler 2011).

Das gezeigte Beispiel unterstellt eine einzelne Beobachtungsgleichung. Es können aber auch *mehrere* Beobachtungsgleichungen zum Einsatz kommen, wenn z. B. das dynamische latente Konstrukt Goodwill über mehrere Indikatoren im Rahmen einer *dynamischen Faktorenanalyse* gemessen wird. Angenommen es gäbe drei Indikatoren für den Goodwill, wobei wie bisher für alle Zeitpunkte $g_t = \gamma g_{t-1} + \beta x_t + \eta_t$ gilt:

$$y_{1t} = \lambda_1 g_t + \varepsilon_{1t}, \quad (2.5)$$

$$y_{2t} = \lambda_2 g_t + \varepsilon_{2t} \text{ und} \quad (2.6)$$

$$y_{3t} = \lambda_3 g_t + \varepsilon_{3t}. \quad (2.7)$$

⁴ Dekimpe *et al.* (2008) präsentieren passende ZRM für eine Vielzahl dynamischer Werbemodelle.

Es folgt das ZRM mit $\mathbf{y}_t = [y_{1t} \ y_{2t} \ y_{3t}]'$, $\mathbf{H} = [\lambda_1 \ \lambda_2 \ \lambda_3]'$ und $\boldsymbol{\varepsilon}_t = [\varepsilon_{1t} \ \varepsilon_{2t} \ \varepsilon_{3t}]'$. Für die Identifizierbarkeit des Modells ist es notwendig, dieses zu restringieren, z. B. indem einer Faktorladung der Wert eins zugewiesen wird (z. B. $\lambda_1 = 1$). Korrelationen der Störterme in den Gleichungen (2.5) bis (2.7) sind über eine volle Kovarianzmatrix \mathbf{R} abzubilden (Petrus *et al.* 2009, S. 138 ff. und 200 ff.).

Der entscheidende Vorteil bei der Nutzung von ZRM für dynamische Werbemodelle ist, dass die latente Größe Goodwill *explizit modelliert* wird und mithilfe des KF auch geschätzt werden kann. Alternative Ansätze verlangen häufig, Modelle durch eine Datentransformation der Schätzung zugänglich zu machen (z. B. mittels Koyck-Transformation) und/oder Carry-Over-Koeffizienten subjektiv festzusetzen bzw. mittels „Grid Search“ zu bestimmen (vgl. hierzu ausführlich die Arbeit von Köhler 2011). Letzteres erschwert inferenzstatistische Aussagen bzgl. der Dynamik, weil keine Standardfehler berechnet werden. Zudem ist die explizite Trennung von Messfehlern ($\boldsymbol{\varepsilon}_t$) und Zufallseinflüssen ($\boldsymbol{\eta}_t$) auf Ebene der dynamischen latenten Größe (\mathbf{a}_t) in ZRM insb. bei Werbemodellen sinnvoll und verhindert Parameterverzerrungen (Naik *et al.* 1998, 2007).

Dieses Beispiel veranschaulicht, dass es verschiedene Möglichkeiten zur Modellierung der Beobachtungsgleichung gibt. Das zweite Beispiel in diesem Abschnitt thematisiert die Flexibilität von ZRM bzgl. der Zustandsgleichung.

Beispiel 2: Univariate Regression mit zeitvariablen Parametern

In vielen Situationen erscheint es unplausibel, Regressionskoeffizienten als intertemporal konstant anzunehmen (Pauwels und Hanssens 2007). Ein einfaches Regressionsmodell mit der abhängigen Variable y_t (z. B. Absatz eines Produktes) und der unabhängigen Variable x_t (z. B. Marketinginvestitionen) und zeitvariablen Parametern lautet:

$$y_t = \alpha_t + \beta_t x_t + \varepsilon_t, \quad \forall t. \quad (2.8)$$

Die Dynamik der beiden Parameter α_t und β_t wird wie folgt modelliert:

$$\alpha_t = \gamma_\alpha \alpha_{t-1} + \tau_\alpha + \eta_{\alpha t} \quad \text{und} \quad (2.9)$$

$$\beta_t = \gamma_\beta \beta_{t-1} + \tau_\beta + \eta_{\beta t}. \quad (2.10)$$

Für die Zustandsraumform gilt dann:

$$\underbrace{\begin{bmatrix} y_t \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} 1 & x_t \end{bmatrix}}_{H_t} \cdot \underbrace{\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}}_{a_t} + \underbrace{\begin{bmatrix} 0 \end{bmatrix}}_d + \underbrace{\begin{bmatrix} \varepsilon_t \end{bmatrix}}_{\varepsilon_t}, \text{ mit} \quad (2.11)$$

$$\underbrace{\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}}_{a_t} = \underbrace{\begin{bmatrix} \gamma_\alpha & 0 \\ 0 & \gamma_\beta \end{bmatrix}}_F \cdot \underbrace{\begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix}}_{a_{t-1}} + \underbrace{\begin{bmatrix} \tau_\alpha \\ \tau_\beta \end{bmatrix}}_c + \underbrace{\begin{bmatrix} \eta_{\alpha t} \\ \eta_{\beta t} \end{bmatrix}}_{\eta_t}, \forall t. \quad (2.12)$$

Fünf Punkte lassen sich zum ZRM in diesem Beispiel anführen:

1. Es kann in einem ZRM durchaus *eine* Beobachtungsgleichung und ein System mit *mehreren* Zustandsgleichungen geben. Die Beobachtung ergibt sich dann aus der (gewichteten) Summe mehrerer dynamischer Zustände.
2. Je nachdem, welche Werte die Parameter $\gamma = [\gamma_\alpha \ \gamma_\beta]'$ und $\tau = [\tau_\alpha \ \tau_\beta]'$ annehmen, resultieren verschiedene dynamische Modelle. Für $\gamma = \mathbf{1}$ und $\tau = \mathbf{0}$ folgen die Parameter einem *Random Walk*. Dieses *nichtstationäre*⁵ dynamische Modell für die Regressionsparameter lässt eine gewisse Variation der Parameter über die Zeit zu, ohne eine spezifische „Richtung“ und „Gestalt“ der Entwicklung vorzugeben (Kim und Nelson 1999). Gilt dagegen $\gamma = \mathbf{1}$ und $\tau \neq \mathbf{0}$, besitzen die Parameter einen *konstanten Trend*. Für $|\gamma| < 1$ folgen die Parameter (unabhängig von τ) einem stationären *AR(1)*-Modell und wird zudem $\tau = (1 - \gamma) \cdot \bar{\tau}$ gesetzt, streben die Parameter immer wieder gegen den „Normalwert“ $\bar{\tau}$, von dem sie nur durch die Zufallsschocks in η_t abweichen (Harvey 1989, S. 410 f.). Ein weiterer Sonderfall liegt für $\gamma = \mathbf{0}$ und $\tau \neq \mathbf{0}$ vor. Hier sind die Parameter nicht mehr intertemporal verknüpft, trotzdem gibt es über η_t *zeitliche* Zufallsschwankungen um den Parametermittelwert τ herum (Lachaab *et al.* 2006, S. 63).
3. Das Modell *kann* dynamisch sein, Dynamik muss aber nicht enthalten sein. Wenn, wie zuvor beim Random Walk, $\gamma = \mathbf{1}$ sowie $\tau = \mathbf{0}$ gilt und wenn zusätzlich alle Elemente in \mathbf{Q}_t (der Kovarianzmatrix von η_t) gleich Null sind, dann sind die beiden Parameter α_t und β_t

⁵ *Stationarität* ist eine wichtige Eigenschaft von Zeitreihendaten. Eine Zeitreihe wird als *stark stationär* bezeichnet, wenn die gemeinsame Verteilung der Beobachtungen zeitunabhängig ist. Ferner wird von *schwacher Stationarität* gesprochen, wenn der Erwartungswert der Zeitreihe konstant und ihre Varianz endlich ist (Greene 2011, S. 923).

intertemporal konstant. Hierdurch reduziert sich das Modell (2.12) zu einem *normalen Regressionsmodell ohne zeitvariable Parameter*.

4. Denkbar ist ebenfalls, dass τ zeitvariabel ist. Folgt τ_t selbst einem dynamischen Modell, dann ist dies im Rahmen eines ZRM zu modellieren, indem der Zustandsvektor *erweitert* wird (Osinga *et al.* 2010):

$$\underbrace{\begin{bmatrix} y_t \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} 1 & x_t & 0 & 0 \end{bmatrix}}_{H_t} \cdot \underbrace{\begin{bmatrix} \alpha_t \\ \beta_t \\ \tau_{\alpha t} \\ \tau_{\beta t} \end{bmatrix}}_{a_t} + \underbrace{\begin{bmatrix} 0 \end{bmatrix}}_d + \underbrace{\begin{bmatrix} \varepsilon_t \end{bmatrix}}_{\varepsilon_t}, \text{ mit} \quad (2.13)$$

$$\underbrace{\begin{bmatrix} \alpha_t \\ \beta_t \\ \tau_{\alpha t} \\ \tau_{\beta t} \end{bmatrix}}_{a_t} = \underbrace{\begin{bmatrix} \gamma_{\alpha} & 0 & 1 & 0 \\ 0 & \gamma_{\beta} & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_F \cdot \underbrace{\begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \\ \tau_{\alpha, t-1} \\ \tau_{\beta, t-1} \end{bmatrix}}_{a_{t-1}} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_c + \underbrace{\begin{bmatrix} \eta_{\alpha t} \\ \eta_{\beta t} \\ \eta_{\tau_{\alpha t}} \\ \eta_{\tau_{\beta t}} \end{bmatrix}}_{\eta_t}, \forall t. \quad (2.14)$$

Durch Nebendiagonalelemente in Q_t ungleich Null können gemeinsame stochastische Zustandsentwicklungen über die Zeit abgebildet werden.

5. Für das Marketing von besonderem Interesse ist ein weiterer Sonderfall des Modells, der sich bei einer Spezifikation von τ_t als Funktion exogener Variablen z_t ergibt (allgemein: $\tau_t = f(z_t)$). Hierdurch lässt sich untersuchen, welchen Einfluss bestimmte Größen auf die *zeitliche Entwicklung* der Regressionsparameter haben (van Heerde *et al.* 2004). Geht man vereinfachend davon aus, dass nur α_t zeitvariabel ist, $f(z_t)$ als lineare Funktion spezifiziert wird und es sich bei den exogenen Größen ebenfalls um die Marketinginvestitionen aus der Beobachtungsgleichung (2.8) handelt ($z_t = x_t$), folgt (Leeftang *et al.* 2009, S. 14):

$$y_t = \alpha_t + \beta x_t + \varepsilon_t, \text{ und} \quad (2.15)$$

$$\alpha_t = \gamma \alpha_{t-1} + \tau x_t + \eta_t, \forall t. \quad (2.16)$$

Dieses Modell mit zeitvariabler „Baseline“ hat bereits die Form eines ZRM und für die nachfolgende Diskussion wird von einem positiven Carry-Over-Koeffizienten ausgegangen ($0 < \gamma \leq 1$).

Praktisch ist nun, dass dieses Modell die Unterscheidung zwischen *Stock-* und *Flow-Variablen* ermöglicht (Hanssens und Dekimpe 2008, S. 509 f.). Typische Flow-Variablen (kumulative Größen) im Marketing sind z. B. Absatz, Umsatz oder Deckungsbeitrag. Diese werden zwischen den Beobachtungszeitpunkten zu einem Wert für diese Periode aufaddiert. Unter Stock-Variablen versteht man dagegen Bestandsgrößen wie z. B. Brand-Equity, Goodwill oder Loyalität. Marketingaktivitäten können nun ganz verschieden auf und vor allem durch diese Größen wirken. Über d_t (hier: βx_t) werden kurzfristige Marketingeffekte berücksichtigt. Sie wirken unmittelbar und nur zu diesem Zeitpunkt über die Messgleichung direkt auf die Flow-Variable („creating lift“). Langfristige Marketingeffekte können über c_t in der Zustandsgleichung abgebildet werden (hier: τx_t). Sie bauen eine Stock-Variable auf („building the base“) und wirken über diese zwar indirekt, aber nachhaltig auf Flow-Variablen.⁶ Bei Hanssens und Dekimpe (2008) handelt es sich bei der Flow-Variablen um den Absatz und so ist die Stock-Variable α_t bspw. als zeitvariable Markenstärke zu interpretieren, die den „Grundabsatz“ bestimmt. Die in der nachfolgenden Tabelle 2.1 dargestellten Szenarien sind nun denkbar.

Tabelle 2.1: Trennung kurz- und langfristiger Marketingeffekte

	$\tau = 0$	$\tau > 0$
$\beta = 0$	Marketing ist <i>wirkungslos</i>	Marketing stärkt nur die <i>Marke</i>
$\beta > 0$	Marketing erzeugt nur <i>Absatz</i>	Marketing erzeugt <i>Absatz</i> und stärkt die <i>Marke</i>

Quelle: Hanssens und Dekimpe (2008, S. 510).

Diese explizite Trennung kurz- und langfristiger Effekte ist eine sehr vorteilhafte Eigenschaft des beschriebenen Modellrahmens und erlaubt eine präzise Ermittlung der *Wirkungsdauer* verschiedener Marketingaktivitäten (Leeflang *et al.* 2009). Sie ist von besonderer Relevanz, wenn es um eine monetäre Evaluierung unterschiedlicher Marketingmaßnahmen geht.

Die Ausführungen zum letzten Beispiel belegen, dass ZRM sehr geeignet sind, zeitvariable Parameter abzubilden. Die große *Flexibilität* des ein-

⁶ Für den Sonderfall $\gamma = 1$ erzeugt die Marketinginvestition aufgrund der nichtstationären Baseline einen *persistenten* Effekt (vgl. z. B. Osinga *et al.* 2010).

heitlichen Modellrahmens, der eine *Vielzahl* möglicher *Parameterdynamiken* *nestet*, die Tatsache, dass *nichtstationäre* Parameterdynamiken handhabbar sind, sowie die Möglichkeit der *Trennung kurz- und langfristiger Marketingeffekte*, können als Hauptgründe angeführt werden, warum derzeitige Publikationen immer seltener andere Modellierungsansätze verwenden (Lee-Flang *et al.* 2009). Zum Beispiel haben die *Piecewise-* und *Moving Window-Regression* den Nachteil, dass sie ineffizient sind, da sie stets nur einen Teil der Daten nutzen.⁷ Ferner führen zu kleine Datenausschnitte zu unrelia- blen Schätzungen, wogegen zu große Datenausschnitte unnötig „grobe“ Ergebnisse liefern und künstliche Autokorrelation in den Ergebnissen erzeugen können (Osinga *et al.* 2010). Parameterdynamik durch Interaktionen mit zeitvariablen Kovariaten zu modellieren ist ebenfalls problematisch, da nur die Variation der Parameter und nicht der zeitliche Verlauf abgebildet wird. Hierbei erfolgt nämlich nur die Modellierung der Parameter als Erwartungswerte in Abhängigkeit von der zeitlichen Variation der exogenen Größen (van Heerde *et al.* 2004, S. 167 f.). Folglich ist z. B. eine explorative Untersuchung von Parameterdynamiken nicht möglich, wenn keine genauen Kenntnisse darüber vorliegen, welche Variablen die zeitliche Variation überhaupt bedingen.

Abschließend sei erwähnt, dass nicht nur autoregressive Prozesse *1. Ordnung* (AR (1)-Modelle) als ZRM formuliert werden können. Viele bekannte Modelle aus der *Zeitreihenanalyse*, wie z. B. autoregressive Modelle (AR (*p*)-Modelle) mit beliebigen Lag-Strukturen, Modelle mit intertemporal korrelierten Störgrößen (MA-Modelle), Modellkombinationen (ARMA-Modelle) sowie Modelle mit exogenen Variablen (ARMAX-Modelle), sind ebenfalls problemlos in die Zustandsraumform überführbar (Shumway und Stoffer 2011, S. 354 ff.). Ein weiteres beliebtes Anwendungsgebiet sind *strukturelle Zeitreihenmodelle*, mit welchen sich Zeitreihen in einzelne Komponenten (Trend, Saison und irreguläre Effekte) zerlegen lassen (Harvey 1989).

⁷ Dieses Vorgehen kann sogar zu einem Interpretationsproblem führen. In Logitmodellen z. B. ist die Skalierung des Fehlerterms zu fixieren, so dass die Parameterwerte implizit von der Varianz des Störterms abhängen. Ein direkter Vergleich von Parameterwerten, deren Schätzungen auf unterschiedlichen Datensätzen basieren, ist dann nicht ohne Weiteres zulässig (Mela *et al.* 1997).

2.3 Zustandsschätzung mit dem Kalman Filter

Im vorherigen Abschnitt wurde gezeigt, dass sehr verschiedene Modelle als ZRM zu formulieren sind und dynamische Zustände von Modell zu Modell unterschiedliche Interpretationen besitzen können. Sobald nun ein Modell in die Zustandsraumform überführt ist, steht mit dem KF (Kalman 1960) ein Werkzeug zur Zustandsschätzung bereit. Der KF ist ein *Rechenalgorithmus*, der rekursiv Schätzwerte $\mathbf{a}_{t|t}$ für den Zustandsvektor \mathbf{a}_t ermittelt. Die Indexnotation „ $t|t$ “ bringt zum Ausdruck, dass die Schätzwerte zum Zeitpunkt t (Wert links neben dem vertikalen Strich) auf den Beobachtungswerten aller Zeitpunkte bis einschließlich t (Wert rechts neben dem vertikalen Strich) basieren. Dabei ist $\mathbf{a}_{t|t}$ die „beste lineare Approximation“ von \mathbf{a}_t und es kann gezeigt werden, dass diese *gefilterten* Zustandswerte optimal im Sinne eines *geringsten mittleren quadratischen Fehlers* sind (Harvey 1989, S. 104 ff.).⁸ Wegen der angenommenen Normalverteilung und Linearität des ZRM ist es mit dem KF möglich, die ersten beiden Momente (Mittelwert und Varianz) des Zustandsvektors \mathbf{a}_t zu bestimmen. Diese lauten:

$$\mathbf{a}_{t|t} = \mathbb{E}[\mathbf{a}_t | \mathbf{Y}_t] \text{ und } \mathbf{P}_{t|t} = \mathbb{E}[(\mathbf{a}_t - \mathbf{a}_{t|t})(\mathbf{a}_t - \mathbf{a}_{t|t})' | \mathbf{Y}_t], \quad (2.17)$$

mit dem Beobachtungsset $\mathbf{Y}_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$. $\mathbf{P}_{t|t}$ ist die Matrix des mittleren quadratischen Fehlers und kann als Schätzungenauigkeit bzw. als Unsicherheit bzgl. $\mathbf{a}_{t|t}$ interpretiert werden (Shumway und Stoffer 2011, S. 325).

Der KF arbeitet folgendermaßen: Sobald eine neue Beobachtung vorliegt, aktualisiert er den Zustandswert mit dieser neu zugänglichen Information. Dabei zeichnen ihn zwei methodische Besonderheiten aus:

1. Der KF wertet wegen der Markov-Eigenschaft von ZRM nicht wieder alle Beobachtungen bis zum Zeitpunkt t neu aus, sondern nur die neueste, sobald diese zur Verfügung steht (Shumway und Stoffer 2011, S. 325). Er ist deshalb ein *effizient* arbeitender Algorithmus. Bei alternativen Verfahren zur Zustandsschätzung werden sehr schnell

⁸ Eng verwandt mit dem Kalman Filter ist der Kalman Smoother. Hierbei werden für die Zustandsschätzwerte zum Zeitpunkt t auch Informationen *zukünftiger* Perioden berücksichtigt (Harvey 1989, S. 154 ff.). Der Kalman Smoother wird im vorliegenden Beitrag aus Platzgründen nicht behandelt. Anwendungen des Kalman Smoothers und entsprechende Erläuterungen sind aber in den nachfolgenden Beiträgen in den Kapiteln 3 und 4 zu finden.

„unhandliche“ Matrix-Dimensionen erreicht, da sämtliche Beobachtungen immer wieder gleichzeitig ausgewertet werden müssen (Kim und Nelson 1999, S. 20 f.).

2. Der KF greift zur Zustandsschätzung nur auf den verwertbaren Teil der neuen Information zu. Der Algorithmus bestimmt dabei zu jedem Zeitpunkt, welcher Anteil der neuen Information die Modellprognose verbessert und welcher nicht. Somit arbeitet der KF *adaptiv*, er berücksichtigt dynamisch die jeweiligen Gegebenheiten und nutzt so den sequentiellen Charakter zeitlich geordneter Beobachtungen (Kim und Nelson 1999, S. 22 f.).

Abbildung 2.2 zeigt die *rekursive Berechnung des KF*. Diese erfolgt in mehreren Schritten, wobei der *Prognose-* und der *Korrekturschritt* den Kern des Algorithmus darstellen (Harvey 1989, S. 105 ff.).

Abbildung 2.2: Schritte zur rekursiven Berechnung des Kalman Filters

1. Wähle Startwerte für die Verteilung des Anfangszustands:

$$\mathbf{a}_{0|0} = \mathbf{a}_0 \text{ und } \mathbf{P}_{0|0} = \mathbf{P}_0. \quad (2.18)$$

2. Prognoseschritt: Prognostiziere die ersten beiden Momente für den Zustand und die Beobachtung, bevor eine neue Beobachtung in t vorliegt:

$$\mathbf{a}_{t|t-1} = \mathbf{F}_t \mathbf{a}_{t-1|t-1} + \mathbf{c}_t, \quad \mathbf{P}_{t|t-1} = \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t' + \mathbf{Q}_t, \quad (2.19)$$

$$\mathbf{y}_{t|t-1} = \mathbf{H}_t \mathbf{a}_{t|t-1} + \mathbf{d}_t, \quad \mathbf{\Sigma}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \mathbf{R}_t. \quad (2.20)$$

3. Berechne das Verhältnis der Prognoseunsicherheiten („Kalman Gain“) und den Prognosefehler mit der neuen Beobachtung in t :

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t' \mathbf{\Sigma}_t^{-1}, \quad \mathbf{e}_t = \mathbf{y}_t - \mathbf{y}_{t|t-1}. \quad (2.21)$$

4. Korrekturschritt: Aktualisiere die ersten beiden Momente der Zustandsprognose mit dem Kalman Gain und dem Prognosefehler:

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{e}_t, \quad \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}. \quad (2.22)$$

5. Wiederhole die Schritte 2 bis 4 für alle $t = 2, 3, \dots, T$ Perioden.

Die Gleichungen des KF können über Standardrechenregeln zur linearen Transformation normalverteilter Größen begründet werden. Herleiten lässt sich insb. der Korrekturschritt mit dem *Satz über die Normal-Korre-*

lation (Greene 2011, S. 1081 f.): Zwei Vektoren \mathbf{v}_1 und \mathbf{v}_2 seien *gemeinsam normalverteilt*:

$$\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \sim N \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} \right). \quad (2.23)$$

Dann gilt für die *bedingte Verteilung* von \mathbf{v}_1 gegeben \mathbf{v}_2 : $\mathbf{v}_1|\mathbf{v}_2 \sim N(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Omega}_{1|2})$ mit $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{v}_2 - \boldsymbol{\mu}_2)$ und $\boldsymbol{\Omega}_{1|2} = \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}$. Mit $\mathbf{v}_1 = \mathbf{a}_t$ und $\mathbf{v}_2 = \mathbf{e}_t$ folgt $\boldsymbol{\mu}_1 = \mathbf{a}_{t|t-1}$ sowie $\boldsymbol{\mu}_2 = \mathbf{0}$ und für die (Ko-)Varianzen gilt $\boldsymbol{\Omega}_{11} = \mathbf{P}_{t|t-1}$, $\boldsymbol{\Omega}_{22} = \boldsymbol{\Sigma}_t$ und $\boldsymbol{\Omega}_{21} = \mathbf{H}_t\mathbf{P}_{t|t-1}$ bzw. $\boldsymbol{\Omega}_{12} = \mathbf{P}_{t|t-1}\mathbf{H}_t'$. Hiermit resultieren unmittelbar die Gleichungen des Korrekturschritts (2.22).

Bevor auf Einzelheiten des KF näher eingegangen wird, ist ein *numerisches Beispiel* zum besseren Verständnis der fünf Schritte des Algorithmus hilfreich.

Beispiel 3: KF-Schätzung eines Random Walk mit konstantem Drift

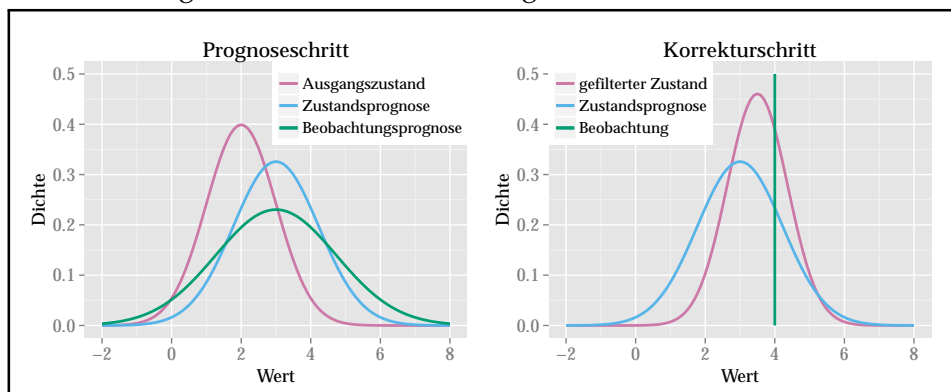
Ausgegangen wird von folgendem ZRM (Petrus *et al.* 2009, S. 35 ff.): $y_t = a_t + \varepsilon_t$, mit $\varepsilon_t \sim N(0, 1,5)$ und $a_t = a_{t-1} + c_t + \eta_t$, mit $c_t = 1$ und $\eta_t \sim N(0, 0,5)$. Für den Ausgangszustand wird zunächst $a_0 = 2$ sowie $P_0 = 1$ gewählt (Schritt 1). Mit diesen Angaben kann der KF für $t = 1$ angewandt werden.⁹

Es folgt zunächst der *Prognoseschritt*. Wegen des exogenen Drifts ist der Mittelwert der Zustandsprognose größer als der des Ausgangszustands, es gilt $a_{1|0} = a_{0|0} + c_1 = 2 + 1 = 3$. Gleichzeitig erhöht sich durch den Störterm η_t die Zustandsunsicherheit $P_{1|0} = P_{0|0} + \sigma_{\eta}^2 = 1 + 0,5 = 1,5$. Für die Beobachtungsprognose folgt $y_1 = a_{1|0} = 3$ und aufgrund der Messungengenauigkeit ε_t erhält man für die dazugehörige Varianz $\Sigma_1 = P_{1|0} + \sigma_{\varepsilon}^2 = 1,5 + 1,5 = 3$. Die *Wahrscheinlichkeitsdichten der Verteilungen* sind in Abbildung 2.3 auf der linken Seite dargestellt.

In $t = 1$ wird nun $y_1 = 4$ beobachtet. Damit beträgt der Prognosefehler im dritten Schritt $e_1 = y_1 - y_1 = 4 - 3 = 1$. Zur Berechnung des Kalman Gains

⁹ Damit die KF-Rekursion durchgeführt werden kann, sind sämtliche Werte unbekannter Parameter im ZRM (z. B. die Varianzen in der Beobachtungs- und Zustandsgleichung) basierend auf Vorkenntnissen anzugeben oder zusätzlich aus den Daten zu schätzen. Der KF ermittelt nur Schätzungen für den Zustandsvektor \mathbf{a}_t . Die Schätzung der unbekannten Parameter wird ausführlich im nächsten Abschnitt 2.4 behandelt.

Abbildung 2.3: Grafische Darstellung einer Kalman Filter-Iteration



Quelle: Eigene Darstellung in Anlehnung an Petris et al. (2009, S. 36).

setzt man die Unsicherheiten der Zustands- und Beobachtungsprognose ins Verhältnis, es resultiert $K_1 = P_{1|0} \cdot \Sigma_1^{-1} = 1,5/3 = 0,5$.

Die Zustandsprognose wird in Schritt Nummer vier, dem *Korrekturschritt*, deshalb im vorliegenden Beispiel mit dem halben Wert des Prognosefehlers korrigiert: $a_{1|1} = a_{1|0} + K_1 \cdot e_1 = 3 + 0,5 \cdot 1 = 3,5$. Die Zustandsunsicherheit ist ebenfalls zu aktualisieren und für sie gilt hier: $P_{1|1} = (1 - K_1) \cdot P_{1|0} = (1 - 0,5) \cdot 1,5 = 0,75$. In Abbildung 2.3 ist der Korrekturschritt rechts zu sehen. Es ist zu erkennen, dass die Korrektur den Mittelwert des Zustands näher an die Beobachtung heranführt. Aufgrund der Beobachtungsunsicherheit ist das Kalman Gain kleiner eins, was bedeutet, dass der neuen Beobachtung nicht vollständig „getraut“ wird. $a_{1|0}$ wird also nicht einfach y_1 gleichgesetzt. Die Zustandsverteilung ist nach der Korrektur *schmäler*, die Varianz ist demnach gesunken. Die neue Beobachtung hat folglich zur Reduktion der Zustandsunsicherheit beigetragen. Diesen Ablauf des KF gilt es nun für alle Zeitpunkte zu wiederholen, mit $a_{1|1}$ und $P_{1|1}$ als Ausgangspunkt für den nächsten Prognoseschritt in $t = 2$.

Wie das Beispiel zeigt, ist das *Kalman Gain* K_t im vorgestellten Algorithmus sehr wichtig. Es bestimmt zu jedem Zeitpunkt t , mit welchem Gewicht der aus der Beobachtung gewonnene Prognosefehler bei der Korrektur von $a_{t|t-1}$ und $P_{t|t-1}$ berücksichtigt wird (Kim und Nelson 1999, S. 24 ff.). Sind die Messungenauigkeiten in der Beobachtungsgleichung gering, sind die Varianzen in R_t , der Kovarianzmatrix der Beobachtungsgleichung, entsprechend klein. In diesem Fall nimmt das Kalman Gain große Werte an, da K_t eine inverse Funktion von R_t ist. Einer neuen Beobachtung wird in diesem Fall viel Vertrauen geschenkt. Sind die Varianzen in

\mathbf{R}_t dagegen groß, so ist das Kalman Gain nur sehr klein (im Extremfall ist $\mathbf{K}_t = 0$ für $\mathbf{R}_t = \infty$), wodurch der Prognosefehler im Korrekturschritt nur gering gewichtet wird. Wegen der Ungenauigkeiten bietet der Prognosefehler nur einen unwesentlichen Informationsgehalt für die Zustandsschätzung. Wenn die Zustandsschätzung sehr sicher ist, sind keine oder wenige zusätzliche Informationen nötig bzw. hilfreich. Daher führt eine große (kleine) Unsicherheit $\mathbf{P}_{t|t-1}$ bzgl. des Schätzwerts zu einer starken (schwachen) Gewichtung der Prognosefehlers. Dieser Punkt verdeutlicht den oben beschriebenen Vorteil der Adaptivität des KF. Der Algorithmus bestimmt zu jedem Zeitpunkt selbst, wie stark die optimale Korrektur der Prognose ausfallen muss und ist hierdurch in der Lage, selbständig und flexibel aus den Daten zu lernen.

Im einfachen univariaten Fall, ohne exogene Variablen und mit konstantem Kalman Gain, folgt als Sonderfall des KF die *exponentielle Glättung 1. Ordnung* für die Beobachtung y_t , wobei hier der Wert für \mathbf{K} (subjektiv) vorzugeben oder aus den Daten zu schätzen ist (Petrís *et al.* 2009, S. 86). Bei der exponentiellen Glättung 1. Ordnung wird allerdings auf ein statistisches Modell verzichtet, weshalb im Gegensatz zum KF keine Aussage zur Unsicherheit bzgl. der Zustandsschätzung getroffen werden kann.

Des Weiteren ist zu erkennen, dass $\mathbf{a}_{t|t}$ in Abhängigkeit vom Vorzeichen von \mathbf{e}_t größer oder kleiner als $\mathbf{a}_{t|t-1}$ ausfällt. Offensichtlich muss $\mathbf{a}_{t|t}$ kleiner als $\mathbf{a}_{t|t-1}$ sein, wenn \mathbf{y}_t kleiner als \mathbf{y}_t ist (et vice versa). Für die Kovarianzmatrix $\mathbf{P}_{t|t}$ gilt dieser Zusammenhang nicht. Immer, wenn das Kalman Gain positiv ist, führt der Korrekturschritt zu einer Reduzierung der Unsicherheit bzgl. des Zustands (Kim und Nelson 1999, S. 24). Das ist verständlich, da – wie zuvor beschrieben – nur dann eine Korrektur stattfindet, wenn die neue Beobachtung einen brauchbaren Informationsgehalt besitzt.

Zustandsprognosen, die mehr als eine Periode in der Zukunft liegen sollen, können ebenfalls mit dem KF berechnet werden. Es ist lediglich der Korrekturschritt (2.22) auszulassen (Harvey 1989, S. 147 ff.). In der oben eingeführten Indexnotation ausgedrückt lauten die Zustandsschätzungen entsprechender Prognosen $\mathbf{a}_{t+1|t-1}$, $\mathbf{a}_{t+2|t-1}$, usw. Mit jeder Periode, in der keine Aktualisierung der Information stattfindet, steigen die Varianzwerte an, da die Unsicherheit wegen der Addition von \mathbf{Q}_t in jedem Filterdurchlauf größer wird, je länger die Nutzung der letzten Beobachtung zurückliegt (vgl. Gleichung (2.19)).

Missing Values sind für den KF unproblematisch, was für Anwendungen

im Marketing sehr vorteilhaft ist (van Heerde *et al.* 2004, S. 167 f.). Hierbei sind zwei verschiedene Situationen zu unterscheiden, sofern \mathbf{y}_t pro Periode mehrere Beobachtungen enthält ($m > 1$).

1. Im Falle *gänzlich fehlender Beobachtungswerte* wird der beschriebene Korrekturschritt (2.22) für relevante Zeitpunkte ausgelassen (Harvey 1989, S. 145 f.). Hierzu setzt man für betreffende Zeitpunkte das Kalman Gain gleich Null. Wenn keine neuen Beobachtungen vorhanden sind, läuft der KF auf Basis der bestehenden Prognosen weiter. Der KF liefert also selbst dann geschätzte Zustandswerte, wenn es für diese Periode(n) keine Beobachtung(en) gibt. Sobald nun aber eine neue Beobachtung vorliegt, wird diese wieder im Korrekturschritt genutzt. Wegen der erhöhten Zustandsunsicherheit erfolgt eine stärkere Gewichtung der neuen Information (vgl. Gleichung (2.21)). Hierdurch sind auch *unregelmäßige Messzeitpunkte* in den Daten problemlos handhabbar, es ist lediglich ein Raster diskreter Zeitpunkte zu finden, auf welches die vorhandenen unregelmäßigen Beobachtungen passen.
2. *Fehlen* allerdings nur *manche* der m Beobachtungen im Vektor \mathbf{y}_t , wäre es nicht sinnvoll, die zu diesem Zeitpunkt vorhandene Information im Beobachtungsvektor ungenutzt zu lassen. Angenommen in \mathbf{y}_t gibt es \tilde{m} vorhandenen Elemente, dann ist es möglich, die Dimensionen des KF derart anzupassen, dass die gesamte vorliegende Information nutzbar ist (Petrís *et al.* 2009, S. 59 f.). Hierzu ist es nötig, eine $m \times m$ Einheitsmatrix zu definieren, aus der die Zeilen zu entfernen sind, die mit den Zeilen der fehlenden Werte in \mathbf{y}_t korrespondieren. Mithilfe dieser $\tilde{m} \times m$ Matrix \mathbf{M}_t wird dann eine neue Beobachtungsgleichung definiert, mit $\tilde{\mathbf{y}}_t = \mathbf{M}_t \mathbf{y}_t$, $\tilde{\mathbf{H}}_t = \mathbf{M}_t \mathbf{H}_t$, $\tilde{\mathbf{d}}_t = \mathbf{M}_t \mathbf{d}_t$, $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{M}_t \boldsymbol{\varepsilon}_t$ und $\tilde{\mathbf{R}}_t = \mathbf{M}_t \mathbf{R}_t \mathbf{M}_t'$.

Letztlich ermöglicht es der direkte Zugang zu den Kovarianzmatrizen $\mathbf{P}_{t|t-1}$, $\boldsymbol{\Sigma}_t$ und $\mathbf{P}_{t|t}$ im KF-Algorithmus, *Konfidenzintervalle* für $\mathbf{a}_{t|t-1}$, \mathbf{y}_t und $\mathbf{a}_{t|t}$ zu bestimmen, um die Güte bzw. Genauigkeit dieser Schätzwerte im Zeitverlauf zu evaluieren. Ein 95 %-Konfidenzintervall für $\mathbf{a}_{t|t-1}$ folgt z. B. aus $\mathbf{a}_{t|t-1} \pm 1,96 \cdot (\text{diag}(\mathbf{P}_{t|t-1}))^{1/2}$ (Petrís *et al.* 2009, S. 63).

2.4 Parameterschätzung mit der Maximum-Likelihood-Methode

Bisher wurde stillschweigend unterstellt, dass sämtliche Parameter des ZRM, die nicht in \mathbf{a}_t enthalten sind und folglich nicht mit dem KF bestimmt werden, bekannt sind.¹⁰ In realen Anwendungen wird dies aber nur sehr selten der Fall sein. So wären z.B. in Beispiel 1 die marginale Wirkung des Werbebudgets (β), der Carry-Over-Koeffizient (γ) sowie die Varianzterme (σ_ε^2 und σ_η^2) zu schätzende konstante Parameter. Es stellt sich daher die Frage, wie die gesuchten Parameter zu ermitteln sind. Im vorliegenden Beitrag wird hierfür die klassische *Maximum-Likelihood(ML)-Methode* vorgestellt und diskutiert, da diese zur Parameterbestimmung lediglich auf der vorgestellten KF-Rekursion fußt.¹¹

Die ML-Methode ist ein sehr beliebter Ansatz zur *Bestimmung unbekannter Parameter* (Greene 2011, S. 549). Die „Likelihood“ ist die Wahrscheinlichkeit des Auftretens beobachteter Daten unter Annahme eines statistischen Modells und konditioniert auf ein bestimmtes Parameterset. Das Ziel ist es, genau die Werte der Parameter zu finden, durch welche die Beobachtungen am „wahrscheinlichsten“ werden. Hierzu wird die Likelihoodfunktion aufgestellt und maximiert. ML-Schätzer besitzen vorteilhafte Eigenschaften: Sie sind *konsistent*, *asymptotisch normalverteilt* und *effizient* sowie *invariant* (Greene 2011, S. 490 ff.). Nachteilig ist, dass die *korrekte Verteilung der Daten* spezifiziert werden muss, um die Likelihoodfunktion aufzustellen.

Bei der Likelihoodfunktion handelt es sich um das Produkt der Dichtefunktionen aller einzelnen Beobachtungen (Greene 2011, S. 553 ff.). Wegen der autoregressiven Form der Zustandsgleichung und der damit fehlenden Unabhängigkeit der Beobachtungen ist es zwingend erforderlich, die Likelihoodfunktion als gemeinsame Dichtefunktion in Größen *bedingter Verteilungen* für die T Beobachtungen von \mathbf{y}_t auszudrücken (Harvey 1989, S. 125 ff.). Mit $\boldsymbol{\theta} \in \Theta$ als $p \times 1$ Vektor der p unbekannten Parameter im Parameterraum Θ folgt (Petrís *et al.* 2009, S. 143 ff.):

¹⁰ Diese konstanten Parameter werden manchmal auch als „Hyperparameter“ bezeichnet (Harvey 1989, S. 103).

¹¹ Für Anwendungen, die zur Parameterschätzung in ZRM die bayes'sche Schätztheorie einsetzen, sei z.B. auf Petrís *et al.* (2009) verwiesen. Ein Beitrag mit Marketingbezug ist bspw. van Heerde *et al.* (2004). Vgl. zur bayes'schen Schätztheorie allgemein auch den Beitrag zu simulationsbasierten Schätzverfahren von Proppe (2009).

$$L(\boldsymbol{\theta}|\mathbf{Y}_t) = \prod_{t=1}^T f(\mathbf{y}_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}). \quad (2.24)$$

Die Konditionierung wird bewusst vertauscht, um zu betonen, dass das eigentliche Interesse an den unbekannten Parametern gegeben der beobachteten Daten besteht (Greene 2011, S. 554). Aus Gründen der Übersichtlichkeit wird außerdem die u. U. notwendige Konditionierung auf unabhängige Variablen (z. B. Marketingvariablen) vernachlässigt. Die gesuchten Parameterwerte erhält man aus der Maximierung der Likelihoodfunktion:

$$\boldsymbol{\theta}_{ML} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} L(\boldsymbol{\theta}|\mathbf{Y}_t). \quad (2.25)$$

Wegen der (jeweiligen) *multivariaten Normalverteilung* der Störgrößen $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$ und des Anfangszustands \mathbf{a}_0 , existiert (2.24) in geschlossener Form. Hierbei ist die Unabhängigkeit der Beobachtungen gegeben der Zustände hilfreich (vgl. Abbildung 2.1). Die Likelihoodfunktion lässt sich nun einfach mittels der *Prognosefehler-Zerlegung* auf Basis der Prognosefehler darstellen (Harvey 1989, S. 126). Sie wird aus schätztechnischen Gründen logarithmiert ($\mathcal{L} = \ln L$) und es folgt (Shumway und Stoffer 2011, S. 335):

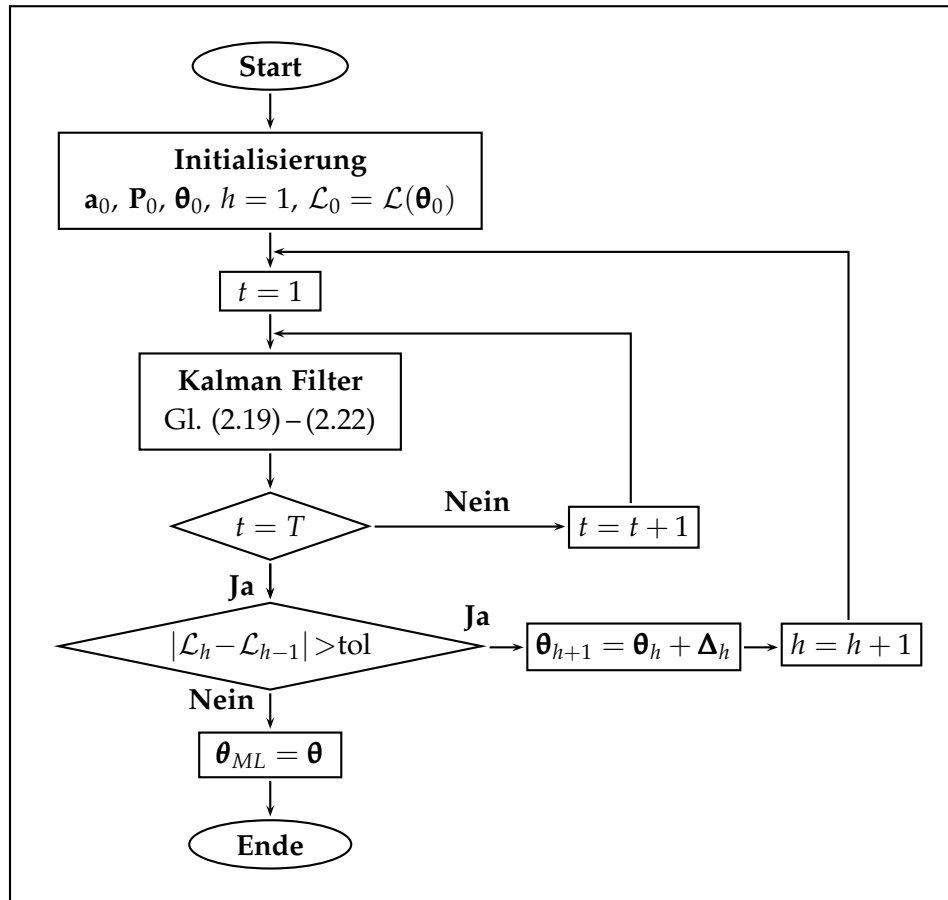
$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}_t) = -0,5 \cdot \sum_{t=1}^T \left[\ln(2\pi^m |\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|) + \mathbf{e}_t(\boldsymbol{\theta})' \boldsymbol{\Sigma}_t(\boldsymbol{\theta})^{-1} \mathbf{e}_t(\boldsymbol{\theta}) \right]. \quad (2.26)$$

Der Faktor 0,5 sowie $2\pi^m$ im Logarithmus-Term von (2.26) dürfen wegen der Unabhängigkeit von $\boldsymbol{\theta}$ vernachlässigt werden. $\boldsymbol{\Sigma}_t$ und \mathbf{e}_t sind dagegen *explizit* Funktionen von $\boldsymbol{\theta}$ und der entscheidende *Vorteil des KF zur Parameterschätzung* ist direkt ersichtlich: Aus der vorgestellten KF-Rekursion (2.18–2.22) geht der Prognosefehler \mathbf{e}_t und dessen Kovarianzmatrix $\boldsymbol{\Sigma}_t$ für jeden Zeitpunkt t unmittelbar hervor. Mit *einem* Durchlauf des KF lässt sich daher die komplizierte und in $\boldsymbol{\theta}$ nichtlineare Log-Likelihoodfunktion als „Nebenprodukt“ direkt konstruieren. Der KF ist daher ein hilfreiches Werkzeug zur Aufstellung der Log-Likelihoodfunktion *jedes dynamischen Modells*, das als ZRM formuliert werden kann.

Aus (2.26) ist außerdem zu erkennen, dass *die* Log-Likelihoodfunktion von Parameterwerten maximiert wird, die möglichst kleine *Prognosefehler* (\mathbf{e}_t)

erzeugen. Die zeitliche Ordnung der Daten spielt im Gegensatz zur einfachen OLS-Regression eine wichtige Rolle, bei der lediglich *Beobachtungsfehler* minimiert werden. Die einzelnen Schritte, die zur Parameterschätzung mit der ML-Methode auf Basis des KF zu durchlaufen sind, stellt Abbildung 2.4 in einem *Flowchart* dar.

Abbildung 2.4: Algorithmus zur Maximum-Likelihood-Schätzung unbekannter (konstanter) Parameter in Zustandsraummodellen



Quelle: Eigene Darstellung in Anlehnung an Kim und Nelson (1999, S. 26).

Weil die Log-Likelihoodfunktion *nichtlinear* in den unbekannten Parametern θ ist, kann das Maximierungsproblem (2.25) nicht geschlossen gelöst werden. Typische *numerische Optimierungsverfahren* sind zu nutzen, um das Maximum der Log-Likelihoodfunktion *iterativ* zu ermitteln (Greene 2011, S. 1132 ff.).

Mit gegebenen Startwerten für die Parameter (θ_0) ist der KF zunächst auf alle Perioden anzuwenden. Dafür werden die zuvor in Abbildung 2.2

beschriebenen Schritte durchgeführt. Nun lässt sich die Log-Likelihoodfunktion gemäß Gleichung (2.26) aufstellen, da der KF für gegebene Parameter θ die Größen Σ_t und e_t für alle Perioden berechnet. Hat sich der Log-Likelihoodwert verbessert, werden für die nächste Optimierungsiteration neue Parameter „ausprobiert“. Beim klassischen *Newton-Raphson-Verfahren* gilt bspw. $\Delta_h = -\mathbf{H}(\theta_h)^{-1} \mathbf{g}(\theta_h)$ für den Updateschritt des Parametervektors, wobei $\mathbf{g}(\theta_h)$ den Gradienten und $\mathbf{H}(\theta_h)$ die Hesse-Matrix der Log-Likelihoodfunktion bezeichnen (Greene 2011, S. 1138). Der iterative Prozess wird so lange wiederholt, bis der Zuwachs des Log-Likelihoodwerts ein zuvor definiertes *Kriterium zur Konvergenz* unterschreitet (z. B. $\text{tol} = 10^{-8}$). Die gesuchten Parameter θ_{ML} sind das Ergebnis des konvergierten Algorithmus.

Die *Hesse-Matrix* der maximierten Log-Likelihoodfunktion bietet zudem den Ausgangspunkt für inferenzstatistische Aussagen bzgl. der Parameter. Die Inverse der negativen Hesse-Matrix ist ein Schätzer für die asymptotische Kovarianzmatrix $\mathbf{V}(\theta) = -\mathbf{H}(\theta)^{-1}$ (Greene 2011, S. 561 ff.). Der maximierte Wert der Log-Likelihoodfunktion liefert ferner die Grundlage zur Modellselektion. Bei *genesteten* Modellen bietet sich z. B. der *Likelihood-Quotienten-Test* an (Greene 2011, S. 564 ff.), *Informationskriterien* (z. B. BIC) eignen sich dagegen bei *nicht-genesteten* Modellen (Greene 2011, S. 573 f.).

Die Log-Likelihoodfunktion von ZRM kann u. U. *multiple lokale Maxima* aufweisen (Petrís et al. 2009, S. 144 f.). Es ist daher ratsam, mehrere Startwerte für den Parametervektor zu testen und θ durch geeignete nichtlineare Transformationen zu restringieren (z. B. mit der Exponentialfunktion für Parameter, die strikt positiv sein müssen). Wegen der Invarianz des ML-Schätzers hat Letzteres keinen Einfluss auf das Maximum der Log-Likelihoodfunktion (Greene 2011, S. 561). Standardfehler für die untransformierten Parameter lassen sich z. B. mithilfe der *Delta-Methode* (Kim und Nelson 1999, S. 13 ff.) oder über einen *simulationsbasierten Ansatz* (Krinsky und Robb 1986, 1990) berechnen.

Bisher nicht angesprochen wurde die zum Start des KF notwendige *Initialisierung* des dynamischen Systems ($\mathbf{a}_{0|0}, \mathbf{P}_{0|0}$). Hier stehen für typische Anwendungen im Marketing prinzipiell *zwei* Möglichkeiten zur Verfügung:

1. Der Anfangszustand $\mathbf{a}_{0|0}$ wird als *unbekannter Parameter* betrachtet und im Vektor θ mitgeschätzt. In diesem Fall gibt es keine Unsicherheit bzgl. dieses Zustands und $\mathbf{P}_{0|0}$ wird auf Null gesetzt (Naik et al. 1998).

2. Es wird ein *fixer Wert* für $\mathbf{a}_{0|0}$ gewählt, bspw. auf Basis sachlogischer Überlegungen oder durch die Schätzung eines einfacheren Modells (van Heerde *et al.* 2004). Da nun Unsicherheit bzgl. des Zustands besteht, wird die Diagonale der Kovarianzmatrix auf einen großen Wert gesetzt ($\mathbf{P}_{0|0} = 10^4 \cdot \mathbf{I}_n$). Hierdurch wird ein hohes Gewicht auf frühe Beobachtungen in der Zeitreihe gelegt.

2.5 Empirische Studie

In der folgenden empirischen Studie wird ein (*bedingtes*) *Multinomiales Logitmodell mit zeitvariablen Parametern* (TVP-MNL-Modell)¹² für *aggregierte Daten* vorgestellt (vgl. zu diskreten Wahlmodellen auch den Beitrag von Temme 2009).¹³ MNL-Modelle sind in der Marketingforschung sehr beliebt, werden i. d. R. aber mit zeitlich invarianten Parametern spezifiziert. Um etwaige Dynamik zu erfassen, kann es aber ratsam sein, das Modell zeitlich flexibel aufzubauen. Ändert sich z. B. die Preissensitivität der Konsumenten, sollte der Entscheider das wissen, um ggf. das Preisniveau anzupassen (Lachaab *et al.* 2006).

Die Basis für das TVP-MNL-Modell bildet die nachfolgende personenspezifische (indirekte) Nutzenfunktion:

$$u_{ijt} = \alpha_{jt} + \beta_t \ln(p_{jt}) + \varepsilon_{jt} + v_{ijt}, \text{ mit} \quad (2.27)$$

$$i = 1, \dots, I, j = 1, \dots, J \text{ und } t = 1, \dots, T.$$

Die zeitvariablen Parameter werden Lachaab *et al.* (2006) folgend als Random Walk-Prozesse modelliert (vgl. das 2. Beispiel aus Abschnitt 2.2):

$$\alpha_{jt} = \alpha_{j,t-1} + \eta_{\alpha_{jt}} \text{ und } \beta_t = \beta_{t-1} + \eta_{\beta_t}, \text{ mit} \quad (2.28)$$

$$j = 1, \dots, J \text{ und } t = 1, \dots, T.$$

Der Nutzen u_{ijt} einer Person i aus der Wahl der Marke j zum Zeitpunkt t kann in zwei Teile zerlegt werden: Den *durchschnittlichen*, über alle I Per-

¹² „TVP“ steht für *time varying parameters*.

¹³ Das hier vorgestellte Vorgehen ist nicht nur auf das MNL-Modell anzuwenden, sondern auf alle Marktanteilsmodelle, die *linearisiert* werden können (Cooper und Nakaniishi 1988).

sonen identischen Nutzen $\delta_{jt} = \alpha_{jt} + \beta_t \ln(p_{jt}) + \varepsilon_{jt}$ und den *personenspezifischen* Teil ν_{ijt} . Dabei ist α_{jt} der markenspezifische und zeitvariable Grundnutzen einer Marke j („Markenwert“) und β_t der zeitvariable Effekt des logarithmierten Preises.¹⁴ Der normalverteilte Störterm ε_{jt} bildet unbeobachtete marken- und zeitspezifische Umstände ab, unter denen die I Personen reagieren, die aber nicht in den Daten erfasst werden, z. B. Regalplatzierung, Couponverfügbarkeit oder Werbung des Herstellers (Besanko *et al.* 1998). Neben J Marken gibt es eine weitere Alternative, die sog. „Nichtwahloption“ ($j = 0$). Zwecks Modellidentifizierung wird der durchschnittliche Nutzen der Nichtwahl auf Null restringiert ($u_{i0t} = \nu_{i0t}$).

Über ν_{ijt} werden personenspezifische Abweichungen von δ_{jt} modelliert. Sind die Konsumenten Nutzenmaximierer und ist ν_{ijt} *unabhängig identisch extremwertverteilt*, dann ist die *Wahrscheinlichkeit*, dass Person i Marke j zum Zeitpunkt t wählt, geschlossen formulierbar. Für die Marktanteile s_{jt} folgt wegen $\exp(\delta_{0t} = 0) = 1$ (Berry 1994, S. 249 f.):

$$s_{jt} = \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^J \exp(\delta_{kt})} \quad \text{und} \quad s_{0t} = \frac{1}{1 + \sum_{k=1}^J \exp(\delta_{kt})}. \quad (2.29)$$

Trifft man eine Annahme für die Abschätzung des Marktpotenzials, ist s_{0t} direkt beobachtbar. Dann lassen sich auf Basis des Zusammenhangs $\ln(s_{jt}/s_{0t}) = \delta_{jt}$ (die Marktanteile werden „invertiert“, Berry 1994) J lineare Schätzgleichungen für den Nutzen der Marken herleiten und diese können in Form eines linearen ZRM aufgestellt werden. Im Rahmen der empirischen Studie werden drei Marken betrachtet ($J = 3$), deswegen resultiert für $t = 1, \dots, T$:

$$\underbrace{\begin{bmatrix} \delta_{1t} \\ \delta_{2t} \\ \delta_{3t} \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \ln(p_{1t}) \\ 0 & 1 & 0 & \ln(p_{2t}) \\ 0 & 0 & 1 & \ln(p_{3t}) \end{bmatrix}}_{\mathbf{H}_t} \cdot \underbrace{\begin{bmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{3t} \\ \beta_t \end{bmatrix}}_{\mathbf{a}_t} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{d}} + \underbrace{\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}}_{\boldsymbol{\varepsilon}_t}, \quad (2.30)$$

¹⁴ Die Spezifikation eines *linearen Preiseffekts* führt bei den hier verwendeten Daten zu einer deutlich *schlechteren Modellanpassung* und wird aus diesem Grund nicht weiter betrachtet.

$$\text{mit } \boldsymbol{\varepsilon}_t \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} \sigma_{\varepsilon_{11}}^2 & \sigma_{\varepsilon_{12}} & \sigma_{\varepsilon_{13}} \\ \sigma_{\varepsilon_{21}} & \sigma_{\varepsilon_{22}}^2 & \sigma_{\varepsilon_{23}} \\ \sigma_{\varepsilon_{31}} & \sigma_{\varepsilon_{32}} & \sigma_{\varepsilon_{33}}^2 \end{bmatrix}}_{\mathbf{R}} \right) \text{ und} \quad (2.31)$$

$$\underbrace{\begin{bmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{3t} \\ \beta_t \end{bmatrix}}_{\mathbf{a}_t} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \cdot \underbrace{\begin{bmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \alpha_{3,t-1} \\ \beta_{t-1} \end{bmatrix}}_{\mathbf{a}_{t-1}} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{c}} + \underbrace{\begin{bmatrix} \eta_{\alpha_1 t} \\ \eta_{\alpha_2 t} \\ \eta_{\alpha_3 t} \\ \eta_{\beta t} \end{bmatrix}}_{\boldsymbol{\eta}_t}, \quad (2.32)$$

$$\text{mit } \boldsymbol{\eta}_t \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} \sigma_{\eta_{\alpha_1}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\eta_{\alpha_2}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\eta_{\alpha_3}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\eta_{\beta}}^2 \end{bmatrix}}_{\mathbf{Q}} \right). \quad (2.33)$$

Weil die unbekannten Parameter im Rahmen der ML-Schätzung nicht restringiert sind, ist es ratsam, diese in den Kovarianzmatrizen \mathbf{R} und \mathbf{Q} adäquat zu transformieren. In \mathbf{Q} reicht es aus, anstatt der Varianzen die Standardabweichungen zu schätzen, um die Nichtnegativitätsbedingung zu erfüllen. Für die volle Kovarianzmatrix \mathbf{R} ist eine Cholesky-Zerlegung $\mathbf{R} = \mathbf{C} \cdot \mathbf{C}'$ praktikabel (Greene 2011, S. 1042 f.). Dabei ist \mathbf{C} eine untere Dreiecksmatrix mit strikt positiv spezifizierten Diagonalelementen, damit sichergestellt wird, dass \mathbf{R} *positiv semidefinit* ist. Die Modellierung einer vollen Kovarianzmatrix \mathbf{R} wird damit begründet, dass aufgrund der Konkurrenzstruktur die Störterme der einzelnen Alternativen *kontemporär korreliert* sein könnten (Besanko *et al.* 1998, S. 1540).

Der KF ermöglicht die Schätzung der zeitvariablen Parameter (α_{1t} , α_{2t} , α_{3t} und β_t). Die Bestimmung der konstanten Parameter in \mathbf{C} und \mathbf{Q} erfolgt mittels der ML-Methode. Beides – die Schätzung der zeitvariablen und der konstanten Parameter – erfolgt jedoch nicht getrennt, sondern einstufig (vgl. Abschnitt 2.4 und dort insb. Tabelle 2.4). Die Elemente α_{10} , α_{20} , α_{30} und β_0 in \mathbf{a}_0 werden ebenfalls als konstante Parameter mitgeschätzt, weil es sich bei den Parameterdynamiken um nichtstationäre Prozesse handelt und keine Vorkenntnisse vorliegen. Hierdurch ist das Modell *genestet* aufgebaut, denn ohne Dynamik in den Parametern (Restriktion: $\boldsymbol{\sigma}_{\boldsymbol{\eta}} = \mathbf{0}$), entsprechen α_{10} , α_{20} , α_{30} und β_0 den Parametern des statischen

MNL-Modells. Das Ergebnis des so mittels der ML-Methode als System geschätzten Modells (*SUR*, vgl. hierzu auch den Beitrag von Schlichthorst 2009) stimmt mit dem der iterierten *FGLS*-Schätzung unter Annahme normalverteilter Fehler überein (Greene 2011, S. 338).

Die verwendeten *Daten* sind im Paket *bayesm* (Rossi 2012) der Open Source Software R (R Core Team 2013) enthalten und stammen aus der Dominick's Fine Foods-Datenbank der Universität Chicago, Booth School of Business.¹⁵ Es werden wöchentliche Absatz- und Preisdaten über 4 Jahre ($T = 208$ Wochen) aus der Produktkategorie Dosenthunfisch verwendet. Die Daten sind über die Geschäfte der Supermarktkette aggregiert und es werden die drei umsatzstärksten Marken *Star Kist* (SK, $j = 1$), *Chicken of the Sea* (CS, $j = 2$) und *Bumble Bee* (BB, $j = 3$) in die Untersuchung einbezogen. Durchschnittlich gibt es in den Geschäften insg. knapp 2 Mio. Kunden pro Woche. Mithilfe dieses Werts wird die (potenzielle) Marktgröße definiert, um die Marktanteile berechnen zu können (Jiang *et al.* 2009). Dabei wird davon ausgegangen, dass ein Kunde bei jedem seiner Einkäufe immer genau eine Dose kauft.

Die ML-Schätzergebnisse für die *konstanten Parameter* sind Tabelle 2.2 zu entnehmen und fallen durchweg sehr zufriedenstellend aus.¹⁶ Alle Parameter besitzen plausible Vorzeichen und sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden.¹⁷ Letzteres gilt insb. auch für $\sigma_{\eta\alpha_j}$ und $\sigma_{\eta\beta}$ und damit ist empirisch belegt, dass zeitliche Variation in den Nutzenparametern vorhanden ist. Ferner ist aus einem Vergleich der markenspezifischen Grundnutzen des MNL-Modells zu erkennen, dass SK den höchsten Markenwert besitzt, gefolgt von CS und BB ($\alpha_{10} > \alpha_{20} > \alpha_{30}$). Für weiterreichende Vergleiche und Interpretationen der Nutzenparameter beider Modelle sei auf die später folgende Diskussion der zeitvariablen Parameter verwiesen.

Wie zuvor beschrieben, ist zur Schätzung der *Kovarianzmatrix* \mathbf{R} die Cholesky-Zerlegung $\mathbf{R} = \mathbf{C} \cdot \mathbf{C}'$ sinnvoll. Zudem erfolgt eine Transformation

¹⁵ Siehe <http://cran.r-project.org/web/packages/bayesm/index.html> für das bereits erwähnte R-Paket *bayesm* (Rossi 2012) sowie für zusätzliche Informationen zu den verwendeten Daten <http://research.chicagobooth.edu/marketing/databases/dominicks/index.aspx>.

¹⁶ Der R-Code zur empirischen Studie ist im Anhang auf S. 44 zu finden.

¹⁷ In diskreten Wahlmodellen lassen sich nur Nutzendifferenzen identifizieren und die α_j -Parameter müssen daher zueinander interpretiert werden. Die negativen Vorzeichen drücken lediglich aus, dass die Nichtwahloption im Vergleich zu den Thunfischmarken sehr häufig gewählt wird.

Tabelle 2.2: Schätzergebnisse für das MNL- und das TVP-MNL-Modell

Parameter	MNL		TVP-MNL	
	Schätzung	Std. Err.	Schätzung	Std. Err.
<i>Nutzenfkt.</i>				
α_{10}^*	-6,016**	0,068	-5,669	0,178
α_{20}	-6,513	0,066	-6,284	0,177
α_{30}	-6,642	0,074	-6,413	0,218
β_0	-4,732	0,211	-6,566	1,093
<i>Störterm</i>				
$\ln(c_{11})$	-0,527	0,050	-0,896	0,057
c_{21}	0,136	0,046	-0,121	0,039
c_{31}	0,131	0,048	-0,122	0,030
$\ln(c_{22})$	-0,486	0,049	-0,719	0,059
c_{32}	0,137	0,042	-0,085	0,033
$\ln(c_{33})$	-0,553	0,049	-1,037	0,076
<i>Dynamik</i>				
$\sigma_{\eta\alpha 1}$			0,050	0,014
$\sigma_{\eta\alpha 2}$			0,048	0,017
$\sigma_{\eta\alpha 3}$			0,082	0,028
$\sigma_{\eta\beta}$			0,507	0,071
\mathcal{L}^{***}	13,371		128,244	

Anmerkung: $j = 1$: Star Kist; $j = 2$: Chicken of the Sea; $j = 3$: Bumble Bee. (*) Im TVP-MNL-Modell beziehen sich die Werte der Parameter α_{j0} , für $j = 1, 2$ und 3 sowie β_0 , auf die Startwerte der zeitvariablen Parameter für $t = 0$. Für das MNL-Modell dagegen handelt es sich um die konstanten Parameterwerte für den gesamten Zeitraum. (**) Fettgedruckte Parameter sind wenigstens auf dem 5 %-Niveau signifikant von Null verschieden (zweiseitig). (***) Positive \mathcal{L} -Werte sind in dieser Anwendung möglich, weil die Log-Likelihoodfunktionen der MNL-Modelle auf Basis der normalverteilten Prognosefehler spezifiziert werden (vgl. Gleichung (2.26)).

der Diagonalelemente in \mathbf{C} , damit diese in der Schätzung stets positiv sind (Log Cholesky-Parametrisierung, vgl. Pinheiro und Bates 1996):

$$\mathbf{C} = \begin{bmatrix} \exp(c_{11}) & 0 & 0 \\ c_{21} & \exp(c_{22}) & 0 \\ c_{31} & c_{32} & \exp(c_{33}) \end{bmatrix}. \quad (2.34)$$

Die geschätzten Werte dieser *Cholesky-Faktoren* sind ebenfalls in Tabelle 2.2 abgetragen, eine direkte Interpretation ist aber schwierig. Aufschlussreicher ist eine nähere Betrachtung der zurücktransformierten Kovarianzmatrix \mathbf{R} selbst, wobei zur Berechnung das simulationsbasierte Verfahren von Krinsky und Robb (1986, 1990) zum Einsatz kommt. Es werden 10.000

Züge aus der (asymptotischen) multivariaten Normalverteilung der relevanten Parameter aus Tabelle 2.2 erzeugt. Für jeden Zug wird \mathbf{C} gebildet und die symmetrische Matrix \mathbf{R} berechnet. Hiernach lassen sich dann die empirischen Mittelwerte sowie Standardabweichungen (die als Standardfehler zu interpretieren sind) über alle 10.000 Züge bilden. Die Ergebnisse dieses Vorgehens lauten für beide Modelle:¹⁸

$$\mathbf{R}^{\text{MNL}} = \begin{bmatrix} \mathbf{0,351} & & \\ (0,035) & & \\ \mathbf{0,081} & \mathbf{0,401} & \\ (0,029) & (0,040) & \\ \mathbf{0,078} & \mathbf{0,103} & \mathbf{0,373} \\ (0,030) & (0,030) & (0,037) \end{bmatrix} \quad \text{und} \quad (2.35)$$

$$\mathbf{R}^{\text{TVP-MNL}} = \begin{bmatrix} \mathbf{0,168} & & \\ (0,020) & & \\ -\mathbf{0,049} & \mathbf{0,255} & \\ (0,016) & (0,029) & \\ -\mathbf{0,050} & -\mathbf{0,027} & \mathbf{0,151} \\ (0,013) & (0,016) & (0,020) \end{bmatrix}. \quad (2.36)$$

Die Varianzterme im TVP-MNL-Modell fallen deutlich kleiner aus als im MNL-Modell. Dies ist nicht weiter verwunderlich, da der Störterm des MNL-Modells auch Variation aufnehmen muss, die sich im TVP-MNL-Modell durch die Parameter-Dynamik erklären lässt. Interessanter sind die Vorzeichen der Kovarianzen, welche im MNL-Modell *positiv* sind, was keinesfalls intuitiv ist. Der Grund für die Verwendung der vollen Kovarianzmatrix war die Vermutung kontemporärer Korrelationen der Wahlalternativen bedingt durch die Konkurrenzstruktur in den Daten (Besanko *et al.* 1998, S. 1540). Demnach sollte z. B. ein *positiver* Nachfrageschock von SK (z. B. eine temporäre Promotionmaßnahme) einen *negativen Effekt* auf CS und BB besitzen und umgekehrt. Die negativen Vorzeichen der Kovarianzterme im TVP-MNL-Modell ergeben folglich inhaltlich mehr Sinn, als die positiven Vorzeichen im MNL-Modell.

Mit Blick auf den vorrangigen Zweck dieser empirischen Studie ist ein Modellvergleich zur Klärung der Frage von Interesse, ob eine Modellspezifikation mit zeitvariablen Parametern sinnvoll und wichtig ist. Zur Beurteilung der internen Validität sind daher in Tabelle 2.3 mehrere *Fitmaße auf Nutzenebene* (linearisierte Ebene des Modells) zusammengefasst.¹⁹

¹⁸ In den Klammern sind die Standardfehler zu finden und für eine bessere Lesbarkeit unterbleibt die Darstellung der oberen Dreiecksmatrizen.

¹⁹ Zum Vergleich: Die maximierten \mathcal{L} -Werte mit linearem Preiseffekt lauten 3,404 (MNL-Modell) bzw. 111,353 (TVP-MNL-Modell). Wie zuvor erwähnt, ist die Spezifikation mit logarithmierten Preisen aufgrund des besseren Modellfits vorzuziehen.

Tabelle 2.3: Fitmaße für das MNL- und das TVP-MNL-Modell

Fitmaße	MNL			TVP-MNL		
	SK	CS	BB	SK	CS	BB
R ² (in %)	42,478	53,715	42,834	61,576	61,213	63,928
MAPE (in %)	11,302	10,603	10,419	8,193	9,859	7,551
RMSE	0,591	0,630	0,606	0,483	0,577	0,481
\mathcal{L}	13,371			128,244		

Anmerkung: Die Ergebnisse des TVP-MNL-Modells basieren auf den Beobachtungsprognosen (siehe Gleichung (2.20)). Für die Fitmaße gilt daher: $R^2 = 100 \cdot \sum_{t=1}^T (y_{t|t-1} - \bar{y})^2 / \sum_{t=1}^T (y_t - \bar{y})^2$; $MAPE = (100/T) \cdot \sum_{t=1}^T |(y_t - y_{t|t-1})/y_t|$; $RMSE = ((1/T) \cdot \sum_{t=1}^T (y_t - y_{t|t-1})^2)^{1/2}$.

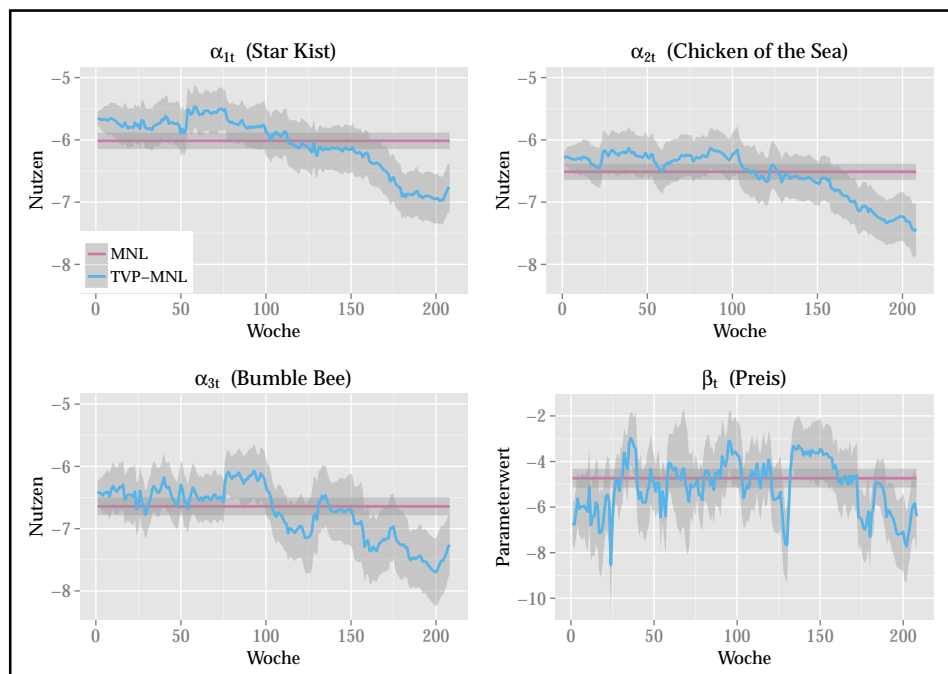
Der *maximierte Wert der Log-Likelihoodfunktion* des TVP-MNL-Modells ist deutlich höher als der des MNL-Modells (128,244 vs. 13,371). Die Signifikanz dieser Differenz ist z. B. mit dem *Likelihood-Quotienten-Test* zu überprüfen ($LR = -2 \cdot (\mathcal{L}_R - \mathcal{L}_U) \sim \chi^2(k)$; Greene 2011, S. 566 f.). Dieser belegt, dass die gemeinsame Nullhypothese intertemporal konstanter Parameter eindeutig verworfen werden muss, da der LR-Wert von 229,747 bei $k = 4$ Freiheitsgraden (vier Varianzterme in \mathbf{Q} sind für das MNL-Modell auf Null zu restringieren) hochsignifikant ist. Auch die Fitmaße für die einzelnen Marken, R^2 , Mean Absolute Percentage Error (MAPE) und Root Mean Square Error (RMSE), stützen einheitlich diesen Befund (vgl. z. B. Barrot 2009). Zwar besitzt bereits das MNL-Modell einen passablen Datenfit, das TVP-MNL-Modell übertrifft diese Werte aber deutlich.

Abbildung 2.5 stellt die *gefilterten* Parameterverläufe beider Modelle grafisch dar. Die 95 %-Konfidenzintervalle belegen, dass sämtliche Parameter zu jedem Zeitpunkt signifikant von Null verschieden sind. Es ist zudem augenscheinlich, dass die zeitvariablen Parameter starken Schwankungen unterliegen und die statischen Parameter des MNL-Modells bestenfalls als eine grobe Näherung nützlich sind.

Die *markenspezifischen Grundnutzen* α_{jt} folgen ca. nach den ersten zwei Jahren einem negativen Trend. Die Verläufe sind aber weder linear, noch stetig und dazu von Marke zu Marke unterschiedlich. Dieser Trend wäre a priori festgelegt nur sehr eingeschränkt modellierbar. Rein deskriptiv lässt sich aus den zeitlichen Entwicklungen der α 's folgern, dass die Grundnutzen der Marken im Vergleich zur Nichtwahloption gesunken sind.²⁰ Diese Entwicklung ist nicht zwingend direkt aus den Marktanteilen (bzw. Absatz-

²⁰ Die negativen Vorzeichen der markenspezifischen Grundnutzen lassen sich durch den hohen Marktanteil der Nichtwahloption erklären (s. o.).

Abbildung 2.5: Grafische Darstellung der zeitlichen Parameterverläufe



Anmerkung: Die 95 %-Konfidenzintervalle sind dunkelgrau hinterlegt eingezeichnet.

werten) zu erkennen, da hier Marketing-Mix-Effekte und andere exogene Einflüsse noch nicht „herausgerechnet“ wurden (Kamakura und Russell 1993). In der Tat sind die Absätze und Marktanteile im Thunfischdatensatz weitestgehend stabil. Brandmanagern bietet das Modell folglich nicht nur wertvolle Erkenntnisse bzgl. der Markenwerte im direkten Konkurrenzvergleich, sondern es ermöglicht zudem, deren zeitliche Entwicklung zu verfolgen. Hierdurch lassen sich frühzeitig Trends, aber auch Probleme erkennen (vgl. hierzu ausführlich Kapitel 4 der vorliegenden Arbeit).

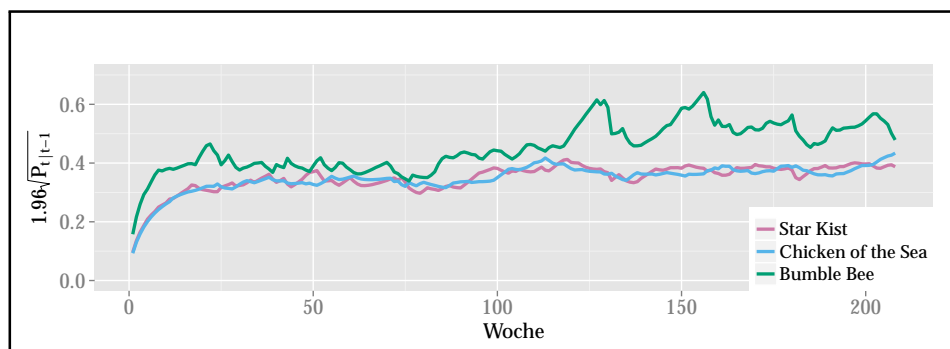
Der Preisparameter β_t (Grafik rechts unten) schwankt im betrachteten Zeitraum ebenfalls beachtlich $[-8,536, -2,979]$ und ist im Betrag oftmals höher als der Preisparameter im MNL-Modell. Der über die Zeit gemittelte Preisparameter des TVP-MNL-Modells hat einen Wert von ca. $-5,100$, der Preisparameter im statischen MNL-Modell hat dagegen einen Wert von $-4,732$. Diese Werte entsprechen hier approximativ Elastizitäten²¹

²¹ Für Preiselastizitäten des (TVP-)MNL-Modells mit logarithmierten Preisen gilt: $\epsilon_{jt} = \beta_{(t)} \cdot (1 - s_{jt})$ (Cooper und Nakanishi 1988). Wegen des vergleichsweise hohen Marktanteils der Nichtwahloption ($s_{0t} > 95\%$) folgt $1 - s_{jt} \approx 1$ und damit entspricht $\beta_{(t)}$ ungefähr der geschätzten Elastizität.

und implizieren, dass Preisentscheidungen auf Basis des MNL-Modells i. d. R. *suboptimal* sind. Um eine Einschätzung zu erhalten, wie hoch die Unterschiede zwischen den optimalen Preisentscheidungen beider Modelle ausfallen, lässt sich die *Amoroso-Robinson-Relation* (Steffenhagen 2008, S. 213) bzw. deren Umformung für *optimale Deckungsbeitragsraten* anwenden ($\text{DB-Rate}^* = (\text{Preis}^* - \text{Stückkosten}) / \text{Preis}^* = 1 / |\text{Elastizität}^*|$).²² Für das TVP-MNL-Modell resultieren optimale Deckungsbeitragsraten zwischen 11,7 % und 33,5 % sowie ein Wert von 19,6 % an der Stelle der mittleren Elastizität. Für das MNL-Modell folgt dagegen 21,1 %. Damit unterscheiden sich die optimalen Preise deutlich und werden auf Basis des MNL-Modells im Durchschnitt zu hoch gesetzt. Storemanager könnten die Ergebnisse des TVP-MNL-Modells demnach einsetzen, um ihre Preisentscheidungen zu verbessern.

In Abbildung 2.6 wird die zeitliche Entwicklung der 95 %-Konfidenzintervalle für die Zustandsprognosen der markenspezifischen Grundnutzen noch einmal genauer verdeutlicht. Zwecks besserer Darstellung sind im Gegensatz zur Abbildung 2.5 alle Grundnutzen auf Null gesetzt. Hierdurch ist nun zu erkennen, wie „sicher“ die Zustandsprognosen auf Basis des KF zu welchem Zeitpunkt ausfallen. Es ist auffällig, dass die Konfidenzintervalle nicht zeitlich konstant sind (vgl. hierzu auch Gleichung (2.19)). Der KF benötigt zudem einige Perioden (ca. 10 bis 20), bis die Konfidenzintervalle ein mehr oder weniger stabiles Niveau erreicht haben, was damit zu begründen ist, dass die Anfangszustände als Parameter mitgeschätzt werden und $P_0 = 0$ gilt.

Abbildung 2.6: Grafische Darstellung der zeitlichen Entwicklung der 95 %-Konfidenzintervalle für die Zustandsprognosen der markenspezifischen Grundnutzen



²² Weil die Kreuzpreiselastizitäten sehr gering sind, unterbleibt eine Berücksichtigung in dieser Beispielrechnung. Vgl. für ein analoges Vorgehen Conley *et al.* (2012, S. 269).

Bei genauerer Betrachtung fällt vor allem das Konfidenzintervall von *BB* auf. In der zweiten Hälfte des Untersuchungszeitraums unterliegt der markenspezifische Grundnutzen deutlichen Schwankungen (vgl. hierzu auch Abbildung 2.5). In diesem Zeitraum vergrößert sich ebenfalls das Konfidenzintervall. Demnach ist nicht nur der *zeitliche Verlauf volatiler*, sondern auch die *Prognoseunsicherheit größer*. Der KF extrahiert folglich wertvolle Einsichten aus den vorhandenen Daten, die bei der Planung weiterer Maßnahmen hilfreich sein können, z. B. durch eine Berücksichtigung in wöchentlichen Preisoptimierungen.

Anhand des hier verwendeten Beispiels wird ein weiterer wichtiger Punkt verdeutlicht: Ist – wie oftmals üblich – eine modellbasierte Entscheidungsunterstützung für *zukünftige* Perioden erwünscht, wäre man hier schlecht beraten, zeitlich konstante Parameter zu verwenden. Abbildung 2.5 belegt, dass die Verzerrungen zum Ende des untersuchten Zeitraums besonders hoch sind. Typischer Kritik (mancher Praktiker) an Einsichten, die auf Vergangenheitsdaten fußen (vgl. hierzu Simon und Fassnacht 2009, S. 136), kann mithilfe des vorgestellten KF begegnet werden, weil dieser adaptiv aus den Daten lernt. Per Definition (s. o.) liefert der KF die besten Prognosen gegeben der *bis dahin* verfügbaren Informationen.

Im Ganzen bleibt festzuhalten, dass die Verwendung zeitvariabler Parameter sinnvoll und wichtig ist. Neben einer besseren Datenanpassung sowie wertvollen Einsichten bzgl. Nutzen- und Preissensitivitätsentwicklungen werden Parameterverzerrungen vermieden. Dies ermöglicht eine korrekte zeitliche Optimierung der Preise und eine brauchbare Entscheidungsunterstützung basierend auf Vergangenheitsdaten.

2.6 Softwareüberblick

Heutzutage gibt es einige Softwarepakete, die eine Schätzung von ZRM mithilfe des KF erlauben. Einen aktuellen und sehr umfassenden Überblick hierzu liefert das Sonderheft (Vol. 41, 2011, <http://www.jstatsoft.org/v41>) des *Journal of Statistical Software* zum Thema „Statistical Software for State Space Methods“ (Commandeur *et al.* 2011).

Die Softwarepakete können grob in zwei Gruppen unterteilt werden.

1. Zum einen gibt es *Standardprogramme* aus dem Bereich der Statistik und Ökonometrie. Es sind insb. *SAS*, *EViews* und *Stata* anzuführen.

Diese Programme besitzen den Vorteil einer sehr einfachen und intuitiven Bedienung. Hierdurch bedingt erlauben sie aber zumeist nur einen eher eingeschränkten Funktionsumfang. Der aufgezeigten Flexibilität der ZRM wird dabei nicht immer hinreichend Rechnung getragen, weil die Software auf typische Zeitreihenmodelle und nicht speziell für Marketinganwendungen ausgerichtet ist.

2. Zum anderen gibt es *moderne matrixbasierte Programme*. Zu dieser Gruppe gehören u. a. *R*, *MATLAB*, *Gauss* oder *Ox*. Diese liefern zwar die notwendige Flexibilität zur Schätzung komplizierter und umfangreicher (Marketing-)Modelle, allerdings zum Preis der weniger einfachen Handhabung. Der Nutzer muss gewillt und fähig sein, z. T. ohne grafische Benutzeroberfläche zu arbeiten und die Methode mittels Matrixalgebra ein- und umzusetzen. Neben einer deutlichen Einarbeitungszeit ist auch die Zeit für die Programmierung eines effizienten und vor allem robust funktionierenden Codes nicht zu unterschätzen. Die Suche und Beseitigung von Fehlern kann in einen mühsamen Trial- und Error-Prozess „ausarten“. Einige Autoren bieten aber auch gut funktionierenden Code auf ihren Homepages an. Zum Beispiel gibt es zu den Lehrbüchern von Kim und Nelson (1999) sowie Shumway und Stoffer (2011) KF-Programme für Gauss bzw. *R*.²³

Neben freiem Code gibt es für einige Programme auch Zusatzpakete, die die KF-Schätzung von ZRM erlauben (z. B. *SsfPack (Basic/Extended)* für *Ox* oder *Gaussx* für *Gauss*). Diese liefern vorgefertigten Code, der erprobt ist und vielfach verwendet wird. Nachteilig ist aber zu nennen, dass diese Programme und Zusatzpakete teilweise *recht teuer zu erwerben* sind. Dieses Hindernis besteht bei der bereits angesprochenen freien und inzwischen sehr weit verbreiteten *Open Source-Software R* nicht.²⁴ Hier gibt es eine Vielzahl von (ebenfalls freien) Zusatzpaketen, die das Arbeiten mit ZRM und dem KF erlauben bzw. erleichtern (vgl. für einen Überblick Tusell 2011). Speziell das Paket *d1m* bietet einen Mittelweg zwischen Bedienungsfreundlichkeit und Funktionsumfang und eignet sich daher für einen *Einstieg* in die Analyse von ZRM mithilfe des KF (Petrís *et al.* 2009). Mit diesem Paket wurden auch beide Modelle der empirischen Studie aus Abschnitt 2.5 ge-

²³ Siehe hierzu <http://www.econ.washington.edu/user/cnelson/markov/prgmlist.htm> für Kim und Nelson (1999) und <http://www.stat.pitt.edu/stoffer/tsa3/> für Shumway und Stoffer (2011).

²⁴ Als weiterer Vorteil kann aufgeführt werden, dass *R* auf Windows-, Linux- und MacOSX-basierten Rechnern zur Verfügung steht.

schätzt (siehe Anhang S. 44). Besonders hervorzuheben ist, dass dlm neben der hier vorgestellten ML-Methode ebenfalls *simulationsbasierte Verfahren* (MCMC, Gibbs-Sampling) zur Bestimmung konstanter Parameter mithilfe des *bayes'schen Ansatzes* bereitstellt und zudem auf einer *numerisch* besonders *stabilen Version* des KF („Square Root Filter“) fußt.

2.7 Zusammenfassende Schlussbetrachtung

Die Berücksichtigung von Dynamik ist im Marketing essentiell. Daher sind adäquate Modelle und Methoden zu nutzen, die Dynamik handhaben können (Leeflang *et al.* 2009). Insbesondere ZRM und der KF eignen sich hervorragend zur Abbildung und Schätzung dynamischer Zusammenhänge. Der vorliegende Beitrag liefert hierzu einen einführenden Überblick und präsentiert Anwendungsbeispiele aus dem Marketing. Es wird veranschaulicht, wie typische dynamische Marketingmodelle in die Zustandsraumform überführt und so dynamische Modellkomponenten mithilfe des KF geschätzt werden können. Mit dem KF kann zudem die Log-Likelihoodfunktion des ZRM gebildet und evaluiert werden. Hierdurch ist es möglich, unbekannte Parameter zu ermitteln, die nicht über den KF selbst bestimmt werden. Anhand einer exemplarischen empirischen Studie wird gezeigt, dass die Berücksichtigung zeitvariabler Parameter nicht nur den Modellfit deutlich verbessert, sondern zudem Verzerrungen in den geschätzten Parametern unterbindet. Ein zunehmendes Softwareangebot zum Einsatz von ZRM und dem KF begünstigt deren weitere Verbreitung.

Abschließend bleibt zu erwähnen, dass zahlreiche *Erweiterungen* des angeführten ZRM existieren, wodurch die *Flexibilität* aber auf Kosten der Komplexität weiter steigt. Naik *et al.* (2008) verwenden z. B. ein ZRM, in dem die Zustandsgleichung *nichtlinear* im Zustand \mathbf{a}_t ist. *Zeitkontinuierliche Zustandsgleichungen* in Form von Differentialgleichungen nutzen bspw. Naik *et al.* (2005). Viele neuere Beiträge, wie z. B. van Heerde *et al.* (2004), setzen anstatt der ML-Methode die *bayes'sche Schätztheorie* zur Bestimmung unbekannter konstanter Parameter ein.

Anhang R-Code für die empirische Studie

```

1 #####
# R-Skript zur Schätzung des MNL- und TVP-MNL-Modells fuer den Tuna-Datensatz. #####
# Hierauf basiert die empirische Studie in Guhl (2014): Zustandsraummodelle und #####
# Kalman Filter im Marketing. R 2.15.0 (oder hoeher) muss installiert sei. #####
# Daniel Guhl: 16.12.2013; d.h.guhl@gmail.com #####
6 #####

# Speicher loeschen und Zielverzeichnis setzen -----
rm(list = ls())
11 setwd("~/Desktop")

# Paketinstallation -----

16 # Paket "mvtnorm" (multivariate Normalverteilung) ggf. installieren und laden
# install.packages("mvtnorm")
require(mvtnorm)

# Paket "dlm" (ZRM + KF) ggf. installieren und laden
21 # install.packages("dlm")
require(dlm)

# Paket "maxLik" (MLE) ggf. installieren und laden
#install.packages("maxLik")
26 require(maxLik)

# Paket "bayesm" (Daten) ggf. installieren und laden
# install.packages("bayesm")
require(bayesm)
31

# Zusatzfunktionen -----
# Zusatzfunktion, um R^2, MAPE und RMSE zu berechnen
Fit <- function (y, yhat) {
36   resi <- y - yhat
   r2 <- 1 - colSums(resi^2) / colSums(t(t(y) - colMeans(y))^2)
   mape <- colMeans(abs(resi / y))
   rmse <- sqrt(colMeans(resi^2))
   list(R2 = 100 * r2, MAPE = 100 * mape, RMSE = rmse)
41 }

# Zusatzfunktion, um die volle Kovarianzmatrix ueber eine Cholesky-Dekomposition zu
# parametrisieren: R = C*C'
Vcovfn <- function (theta) {
46   nr <- .5 * (sqrt(1 + 8 * length(theta)) - 1)
   C <- matrix(0, nr, nr)
   C[lower.tri(C, diag = TRUE)] <- theta
   # strikt positive Diagonalelemente, damit R positiv semidefinit ist
   diag(C) <- exp(diag(C))
51   return(tcrossprod(C))
}

# Datenaufbereitung -----
56

# Bsp. aus Jiang et al. (2009), 142 f.; 3 Thunfisch-Produkte/Marken:
# Star Kist (SK), Chicken of the Sea (CS), Bumble Bee (BB)
data(tuna) # Thunfisch-Daten laden

61 # Anzahl der verwendeten Wochen festlegen; hier 208
nT <- 208
tuna <- tuna[tuna$WEEK <= nT, ]

# Anzahl der Konsumenten pro Woche
66 customers <- tuna$FULLCUST
# Thunfischabsatz pro Woche fuer die drei Marken
units <- tuna[, c(2, 3, 5)]
# logarithmierte Preise pro Woche fuer die drei Marken
lnprice <- tuna[, c(16, 17, 19)]
71

# Marktanteile (mengenmaessig) fuer die 3 Marken berechnen
# Marktgroesse = Mittelwert der Geschaeftsbesucher, vgl. Jiang et al. (2009)
share <- units / mean(customers)

```

```

76 # Marktanteil der Nichtwahloption
shareOG <- 1 - rowSums(share)

# delta-Matrix der 3 Marken
ym <- as.matrix(log(share / shareOG))
# (log) Preis-Matrix der 3 Marken
81 pm <- as.matrix(lnprice)

# Die Dosen besitzen unterschiedliche Groessen (6 oz und 6.12 oz). Damit die Preise ver -
# gleichbar sind, wird der Preis fuer CS auf eine 6.12 oz-Dose umgerechnet. Star Kist hat
# entgegen der Angaben im bayesm-Paket eine 6.12 oz-Dose!
86 # http://research.chicagobooth.edu/marketing/databases/dominicks/categories/tna.aspx
pm[, 2] <- log(exp(pm[, 2]) * 6.12 / 6)

# ZRM + LL-Fkt definieren -----

91 # Funktion, um die relevanten Matrizen des multivar. ZRMs zu spezifizieren
# Achtung: dlm verwendet eine eigene Notation
BuildTVPMMNL <- function (theta) {
  model <- list(
96     m0 = theta[11:14], # a0
     CO = diag(0, 4), # P0
     FF = cbind(diag(3), 0), # Ht
     GG = diag(4), # Ft
101     V = Vcovfn(theta[1:6]), # R = C*C'
     W = diag(theta[7:10]^2, 4), # Q
     # Stellen mit exogenen Variablen in FF (also Ht) definieren
     JFF = cbind(diag(0, 3), c(1, 2, 3)),
     X = pm) # exogene Variablen (Preise) in Ht
  return(model)
106 }

# LogLikelihood-Funktion ermitteln
LLfn <- function (theta) {
  mod <- BuildTVPMMNL(theta)
111  ll <- dlmLL(ym, mod)
  return(-ll)
}

# Schaetzung des TVP-MNL-Modells -----

116 # Startwert fuer die konstanten Parameter
theta0 <- c(-0.5, 0, 0, -0.5, 0, -0.5, # C bzw R
           0.05, 0.05, 0.05, 0.1, # Q
           -6.1, -6.5, -6.6, -4.7) # x0

121 # Parameternamen definieren
namestheta0 <- c("ln(c_11)", "c_21", "c_31", "ln(c_22)", "c_32", "ln(c_33)",
               "sig_eta_a1", "sig_eta_a2", "sig_eta_a3", "sig_eta_b",
126               "alpha_10", "alpha_20", "alpha_30", "beta_0")

names(theta0) <- namestheta0

# MLE TVPMMNL
restTVPMMNL <- maxLik(LLfn, start = theta0, method = "BFGS",
131                      print.level = 1, tol = 1e-12, iterlim = 200)

# Ergebnis-Tabelle
summary(restTVPMMNL)

136 # Tabelle als .txt-Datei abspeichern
write.table(round(summary(restTVPMMNL)$estimate, 3), "Res.TVPMMNL.txt", dec = ".", sep = " ")

# Optimierungs-Check: Sind alle Eigenwerte der Hessematrix positiv?
min(eigen(-hessian(restTVPMMNL))$values) # ja
141

# Simulation nach Krinsky und Robb (1986, 1990) zur Berechnung der ruecktransformierten
# Kovarianzmatrix R
nS <- 10000L # Anzahl der Zufallsziehungen
set.seed(123456789) # Feste Seed-Nummer fuer Reproduzierbarkeit
146 # Ziehung der Zufallszahlen aus der asymptotischen Verteilung der Parameter
msim.TVPMMNL <- rmvnorm(nS, coef(restTVPMMNL)[1:6], vcov(restTVPMMNL)[1:6, 1:6])

# Leeres Array fuer die Ergebnisse
arres.TVPMMNL <- array(NA, c(3, 3, nS))
151 for (ix in 1:nS) {
  arres.TVPMMNL[, , ix] <- Vcovfn(msim.TVPMMNL[ix, ])

```

```

    }

    # Mittelwert und Std-Fehler
156 arres.TVPMNL.mean <- c(apply(arres.TVPMNL, c(1, 2), mean))[c(1, 2, 3, 5, 6, 9)]
    arres.TVPMNL.sd <- c(apply(arres.TVPMNL, c(1, 2), sd))[c(1, 2, 3, 5, 6, 9)]

    # Ergebnis-Tabelle
161 resTVPMNLR <- cbind(arres.TVPMNL.mean, arres.TVPMNL.sd,
                      arres.TVPMNL.mean / arres.TVPMNL.sd,
                      2 * (1 - pnorm(abs(arres.TVPMNL.mean / arres.TVPMNL.sd))))
    rownames(resTVPMNLR) <- c("r_11", "r_21", "r_31", "r_22", "r_32", "r_33")
    colnames(resTVPMNLR) <- c("Estimate", "Std. error", "t value", "Pr(> t)")
    round(resTVPMNLR, 3)

166 # Tabelle als .txt-Datei abspeichern
    write.table(round(resTVPMNLR, 3), "Res.TVPMNL_R.txt", dec = ".", sep = " ")

    # Zustandsschaetzung auf Basis der ML-Ergebnisse der konstanten Parameter
171 filtTVPMNL <- dlmFilter(ym, BuildTVPMNL(coef(resTVPMNL))) # TVP-MNL-Modell
    filtTVPMNL$resTVPMNL <- resTVPMNL

    # KF-Ergebnisse speichern
176 save(filtTVPMNL, file = "Res.KF_TVPMNL.Rdata", compress = TRUE, compression_level = 9)

    # Schaetzung des MNL-Modells -----

    # MLE MNL: Achtung "genau" Null fuer die "sig_eta-Parameter" fuehrt zu numerischen Pro-
181 # blemen im dlm-Paket. Die mit 1e-18 erzielten Ergebnisse entsprechen aber de facto denen
    # einer ML-Schaetzung des SUR-Modells. Das Vorgehen wird an dieser Stelle nur der Ein-
    # fachheit halber gewaehlt.

    # Startwert fuer die konstanten Parameter
186 theta0 <- c(-0.5, 0, 0, -0.5, 0, -0.5, # C bzw R
              1e-18, 1e-18, 1e-18, 1e-18, # Q
              -6.1, -6.5, -6.6, -4.7) # x0
    names(theta0) <- namestheta0

191 # MLE MNL
    resMNL <- maxLik(LLfn, start = theta0, fixed = c(7, 8, 9, 10), method = "BFGS",
                    print.level = 1, tol = 1e-12, iterlim = 200)

    # Ergebnis-Tabelle
196 summary(resMNL)

    # Tabelle als .txt-Datei abspeichern
    write.table(round(summary(resMNL)$estimate, 3), "Res.MNL.txt", dec = ".", sep = " ")

201 # Optimierungs-Check: Sind alle Eigenwerte der Hessematrix positiv?
    min(eigen(-hessian(resMNL)[-c(7:10), -c(7:10)])$values) # ja

    # Simulation nach Krinsky und Robb (1986, 1990) zur Berechnung der ruecktransformierten
    # Kovarianzmatrix R
206 set.seed(123456789) # Feste Seed-Nummer fuer Reproduzierbarkeit
    # Ziehung der Zufallszahlen aus der asymptotischen Verteilung der Parameter
    msim.MNL <- rmvnorm(nS, coef(resMNL)[1:6], vcov(resMNL)[1:6, 1:6])

    # Leeres Array fuer die Ergebnisse
211 arres.MNL <- array(NA, c(3, 3, nS))
    for (ix in 1:nS) {
        arres.MNL[, , ix] <- Vcovfn(msim.MNL[ix,])
    }

216 # Mittelwert und Std-Fehler
    arres.MNL.mean <- c(apply(arres.MNL, c(1, 2), mean))[c(1, 2, 3, 5, 6, 9)]
    arres.MNL.sd <- c(apply(arres.MNL, c(1, 2), sd))[c(1, 2, 3, 5, 6, 9)]

    # Ergebnis-Tabelle
221 resMNLr <- cbind(arres.MNL.mean, arres.MNL.sd,
                  arres.MNL.mean / arres.MNL.sd,
                  2 * (1 - pnorm(abs(arres.MNL.mean / arres.MNL.sd))))
    rownames(resMNLr) <- c("r_11", "r_21", "r_31", "r_22", "r_32", "r_33")
    colnames(resMNLr) <- c("Estimate", "Std. error", "t value", "Pr(> t)")
    round(resMNLr, 3)

226 # Tabelle als .txt-Datei abspeichern
    write.table(round(resMNLr, 3), "Res.MNL_R.txt", dec = ".", sep = " ")

```

```

231 # Zustandsschaetzung auf Basis der ML-Ergebnisse der konstanten Parameter
    filtMNL <- dlmFilter(ym, BuildTVPMNL(coef(resMNL))) # MNL-Modell
    filtMNL$resMNL <- resMNL

    # KF-Ergebnisse speichern
236 save(filtMNL, file = "Res.KF_MNL.Rdata", compress = TRUE, compression_level = 9)

    # Modellvergleich -----

241 # R^2, MAPE und RMSE berechnen (basierend auf 1-step-ahead-Prognosen)
    fit.TVPMNL <- Fit(filtTVPMNL$y, filtTVPMNL$f)
    fit.MNL <- Fit(filtMNL$y, filtMNL$f)

    # Ergebnis-Tabelle
246 fitmat <- matrix(c(fit.MNL$R2, fit.TVPMNL$R2, fit.MNL$MAPE, fit.TVPMNL$MAPE,
                     fit.MNL$RMSE, fit.TVPMNL$RMSE), 3, 6, byrow = TRUE)
    rownames(fitmat) <- c("R^2", "MAPE", "RMSE")
    colnames(fitmat) <- c("SK(MNL)", "CotS(MNL)", "BB(MNL)",
                        "SK(TVP-MNL)", "CotS(TVP-MNL)", "BB(TVP-MNL)")
251 round(fitmat, 3)

    # Tabelle als .txt-Datei abspeichern
    write.table(round(fitmat, 3), "Fit.txt", dec = ".", sep = " ")

256 # LR-Test
    L.Ratio <- - 2 * (logLik(resMNL) - logLik(resTVPMNL))
    p.val <- 1 - pchisq(L.Ratio, 4) # df. = 4, da 4 Varianzpar. in Q auf 0 gesetzt werden

    # Ergebnis-Tabelle
261 LL.RT <- cbind(logLik(resMNL), logLik(resTVPMNL), L.Ratio, p.val)
    colnames(LL.RT) <- c("LL.MNL", "LL.TVPMNL", "L.Ratio", "p.val")
    round(LL.RT, 3)

    # Tabelle als .txt-Datei abspeichern
266 write.table(round(LL.RT, 3), "LL.RT.txt", dec = ".", sep = " ")

# EDNE #####

```


Kapitel 3

Endogenität in Logitmodellen mit zeitvariablen Parametern

„Failure to account for [...] endogeneity in the deterministic components of utility has the potential to bias the parameter estimates of the marketing-mix variables.“
Villas-Boas und Winer (1999, S. 1324)

3.1 Einleitung

Zuletzt haben *Logitmodelle für aggregierte Daten* in der Marketingforschung stetig an Popularität gewonnen (Besanko *et al.* 1998; Dubé *et al.* 2002; Park und Gupta 2009). Sie gehören heutzutage zu den wichtigsten empirischen Werkzeugen, um Märkte mit differenzierten Produkten zu modellieren und zu untersuchen (Ackerberg *et al.* 2007; Reiss und Wolak 2007). Drei Gründe lassen sich für die Popularität anführen: 1.) Die Modelle fußen auf der *ökonomischen Theorie der Nutzenmaximierung*, 2.) sie können sowohl *Endogenität* einzelner Variablen als auch *Konsumentenheterogenität* handhaben¹ und 3.) besitzen „geringe“ *Datenanforderungen* als disaggregierte Modelle.

¹ Shugan merkt bereits vor 10 Jahren an, *Heterogenität* und *Endogenität* haben in der Nachfragemodellierung im Marketing in einem Maße Relevanz erreicht, dass Gutachter hochrangiger internationaler Journale die Berücksichtigung dieser Phänomene für *obligatorisch* halten und sie auf ihre „Reviewer-Checklisten“ aufgenommen haben (Shugan 2004).

Im Einzelnen:

1. Die *theoretische Fundierung* erlaubt es zum einen, *strukturelle Parameter*² zu schätzen, die gut zu *interpretieren* sind und eine *ökonomische Bedeutung* besitzen (Dubé *et al.* 2002), und zum anderen, komplexe Konkurrenzstrukturen sehr *sparsam parametrisiert* abzubilden (Akerberg *et al.* 2007).
2. Inzwischen belegen viele Quellen in der einschlägigen Literatur, dass *Endogenität* (vor allem Preisendogenität) in nichtexperimentellen Daten ein ernsthaftes Problem darstellt. Von Endogenität wird gesprochen, wenn erklärende Variablen mit dem Störterm des Modells korrelieren (Greene 2011, S. 219 ff.) und diese Endogenität verursacht verzerrte Schätzergebnisse, was zu falschen Implikationen führen kann (Villas-Boas und Winer 1999; Chintagunta 2001; Chintagunta *et al.* 2005). Deshalb ist es für viele Fragestellungen unabdingbar, entsprechende Ansätze zu nutzen, die dieses Problem handhaben können (Berry 1994). Evident wichtig ist insb. im Marketing die Möglichkeit zur Berücksichtigung der *Konsumentenheterogenität*, da unterschiedliche Präferenzen und Bedürfnisse *einzelner* Konsumenten das Fundament des Marketing bilden (Allenby und Rossi 1999). Die Verwendung homogener Modelle kann ebenfalls zu verzerrten Ergebnissen führen (Chintagunta *et al.* 1991). In der modernen empirischen Marketingforschung und -praxis ist die Modellierung der Heterogenität daher de facto der Standard.
3. Ein letzter Punkt, der für die Popularität eingangs genannter Modelle spricht, ist die Tatsache, dass lediglich *aggregierte Daten* notwendig sind. Dies ist ein großer Vorteil, da solche Daten für Unternehmen leichter bzw. kostengünstiger zugänglich und handhabbar sind (Bodapati und Gupta 2004) und die Kritik zweifelhafter Repräsentativität weniger akut ist, als bei disaggregierten Daten (z. B. Haushaltspanels) (Gupta *et al.* 1996).

Neben offenkundigen Vorzügen der angeführten Logitmodelle für aggregierte Daten vernachlässigen diese aber i. d. R. eine *dynamische Flexibilität*. Es wird unterstellt, dass sich die Nutzenstrukturen der Konsumenten über die Zeit *nicht* verändern. Gleichwohl sprechen aber viele Gründe dafür, warum genau das der Fall sein könnte. Exemplarisch zu nennen sind (Lee-

² „Strukturelle“ Parameter betreffen die Nutzenbildung des Konsumenten und haben damit einen direkten Bezug zu dessen Nutzenmaximierungsproblem.

flang *et al.* 2009): *Saisonalität und Trends im Markt, Produktinnovationen und eine variierende Produktvielfalt, Auf- und Abbau psychischer Größen* (z. B. Markenbekanntheit, Einstellungen, usw.) sowie *Loyalität und Abwechslungsbedürfnis* der Konsumenten. Angesichts der heutzutage verfügbaren Datensätze mit *Längsschnittdimensionen*, die sich teilweise über *mehrere Jahre* erstrecken, erscheint die Annahme vollkommen „stabiler“ Nutzenstrukturen fragwürdig. Beispielsweise kommen bei Park und Gupta (2012) Quartalsdaten von 1981 bis 1998 zum Einsatz, Jiang *et al.* (2009) nutzen Daten über eine Zeitspanne von fast 400 Wochen und Chintagunta *et al.* (2002) verwendet 196 Wochen.

In diskreten Wahlmodellen für *disaggregierte Daten* werden dynamische Effekte häufig über *Kauffeedbackvariablen* oder *Referenzpreise* berücksichtigt. Kauffeedbackvariablen basieren auf individuellen vergangenen Wahlentscheidungen (Guadagni und Little 1983; Keane 1997). Referenzpreise dagegen basieren auf den individuellen Preishistorien der Konsumenten (Winer 1986; Mazumdar *et al.* 2005; Moon *et al.* 2006). Sofern aber nur aggregierte Daten vorliegen, ist die Verwendung von Kauffeedbackvariablen und/oder Referenzpreisen nicht umsetzbar, da die individuellen Informationen fehlen. Teilweise werden einzelne Größen des Wahlmodells auch durch eine *Reparametrisierung* als Funktionen zeitvariabler Kovariate modelliert (Jedidi *et al.* 1999). Dieses Vorgehen ist ebenfalls sehr beliebt und auch mit aggregierten Daten einfach anzuwenden (Foekens *et al.* 1999; Kopalle *et al.* 1999). Hierdurch wird aber nur die *zeitliche Variation* der dynamischen Größen (als bedingter Erwartungswert) modelliert und nicht deren *zeitliche Entwicklung* abgebildet (van Heerde *et al.* 2004, S. 167). Schließlich gibt es die Möglichkeit, die Daten zu zerteilen und z. B. *Vorher-Nachher-Analysen* durchzuführen. Dazu ist anzumerken, dass absolute Parameterwerte in Logitmodellen wegen Skalierungsunterschieden nicht über mehrere Datenbasen verglichen werden dürfen (Mela *et al.* 1997). Damit reduziert sich die Nützlichkeit des besagten Vorgehens auf vergleichbare Maße, z. B. marginale Effekte oder Elastizitäten, und die Zerteilung führt stets zu einer Reduktion der statistischen Effizienz, da immer nur ein Teil der Daten zur Analyse genutzt werden kann.

Die genannten Vorgehensweisen – sofern anwendbar – sind restriktiv, sie erlauben Dynamik nur in einer eingeschränkten Form und es sind a priori zusätzliche Annahmen (z. B. bzgl. erklärender Variablen oder potenzieller Verläufe) notwendig. Im vorliegenden Beitrag wird daher ein *alternativer Weg* eingeschlagen: Hier werden *Zustandsraummodelle* (ZRM) und der *Kalman Filter* (KF) (Kalman 1960) genutzt, die eine *flexible* Modellierung und

Schätzung der Dynamik in Wahlmodellen ermöglichen. Die Parameter der Nutzenfunktionen dürfen stochastisch über die Zeit variieren und werden als Random Walk-Prozesse spezifiziert (Lachaab *et al.* 2006). In aktuellen Beiträgen zur „Marketingdynamik“ wird dieser Modellierungs- und Schätzansatz aufgrund vieler günstiger Eigenschaften zunehmend eingesetzt (vgl. Leeftang *et al.* 2009 und Kapitel 2 der vorliegenden Arbeit), i. d. R. allerdings nicht zusammen mit Logitmodellen.

Sriram *et al.* (2006) sowie Sriram und Kalwani (2007) nutzen als Erste ebenfalls ZRM, um Dynamik in Logitmodellen für aggregierte Daten zuzulassen. Im Gegensatz zu diesen Quellen geht der vorliegende Beitrag aber nicht nur davon aus, dass die Markenpräferenzen der Konsumenten zeitvariabel sein dürfen, sondern es wird zugelassen, dass auch die Marketing-Mix-Sensitivitäten der Konsumenten über die Zeit variieren, was in vielen Situationen sehr wichtig ist (vgl. allgemein Lachaab *et al.* 2006). Hierzu gibt es zahlreiche empirische Befunde: Marketing-Mix-Sensitivitäten ändern sich bspw. aufgrund saisonaler Einflüsse (Chevalier *et al.* 2003; Meza und Sudhir 2006). Zudem ist es denkbar, dass häufige Preispromotionaktivitäten des Handels dazu führen, dass die Preissensitivität der Konsumenten steigt (Jedidi *et al.* 1999; Ataman *et al.* 2010) und Erdem *et al.* (2008) zeigen, dass auch Werbung einen Einfluss auf die Preissensitivität besitzt.

Sofern endogene Variablen einen zeitvariablen Effekt besitzen, kommt es nun aber zu einem bedeutsamen *methodischen Problem*, Endogenität zu beseitigen und unverzerrte Ergebnisse zu erhalten. In einer Reihe von Publikationen belegt Kim, dass Standardverfahren nicht zulässig sind, wenn endogene Variablen einen zeitvariablen Effekt besitzen (Kim 2006, 2010; Kim und Kim 2011). Der von Sriram *et al.* (2006) und Sriram und Kalwani (2007) eingesetzte Schätzer, welcher auf der generalisierten Momentenmethode (GMM) basiert, ist daher im hier betrachteten Fall *nicht zulässig*. Kim schlägt deshalb alternativ einen Ansatz mit auf sog. *Kontrollfunktionen* (CF) vor. Hierbei wird weiterhin die endogene Variable in ihrer ursprünglichen Form verwendet, zusätzlich werden aber weitere Variablen (die Kontrollfunktionen) im Modell ergänzt, mit deren Hilfe die Endogenitätsverzerrung behoben wird. Zudem erlaubt der CF-Ansatz das direkte Testen eines etwaigen Endogenitätsproblems. Aus den genannten Gründen wird diese von Kim vorgeschlagene Methode auch im vorliegenden Beitrag verwendet und erstmalig auf heterogene Logitmodelle für aggregierte Daten übertragen.

Der vorliegende Beitrag beschäftigt sich mit den folgenden Forschungsfragen:

1. Ist es schätztechnisch möglich, zeitvariable Parameter in Logitmodellen für aggregierte Daten unter Berücksichtigung von Preisendogenität und Heterogenität likelihoodbasiert zu bestimmen?
2. Welche Konsequenzen hat es, wenn Endogenität falsch berücksichtigt wird?
3. Was passiert, wenn Endogenität ignoriert wird?

Diese Fragen werden mithilfe einer *Simulationsstudie* untersucht und anhand einer *empirischen Studie* veranschaulicht. Es wird dabei das Ziel verfolgt, Logitmodelle für aggregierte Daten besser zu verstehen. Angesichts der eingangs erwähnten Popularität und Relevanz dieser Modelle ist es wichtig, zu wissen, wie diese Modelle im Falle von Dynamik zu spezifizieren sind. Neben einer ökonometrischen Motivation für die genannten Forschungsfragen sind die Antworten von großer Relevanz für die Marketingforschung und -praxis, da Dynamik gerade hier bedeutsam ist (Lee-flang *et al.* 2009).

Dem vorliegenden Beitrag liegt folgender *Aufbau* zugrunde: Zunächst wird in Abschnitt 3.2 das Modell vorgestellt und erörtert. Ebenfalls wird dargestellt, wie der CF-Ansatz arbeitet. Hiernach wird im Abschnitt 3.3 beschrieben, wie konstante sowie zeitvariable Modellparameter zu schätzen sind und wie auf Endogenität getestet werden kann. Im Abschnitt 3.4 wird mithilfe einer Simulationsstudie untersucht, welche Leistungsfähigkeit der postulierte Ansatz besitzt, was im folgenden Abschnitt 3.5 auch anhand einer empirischen Studie illustriert wird. Der Beitrag schließt mit Abschnitt 3.6, welcher die Ergebnisse zusammenfasst und Anregungen für künftige Forschungsarbeiten liefert.

3.2 Modellierung

Das Modell besteht aus zwei Teilen, dem *Nachfrage-* und dem *Preismodell*. Das Nachfragemodell basiert auf dem heterogenen („mixed“) Logitmodell (MXL) für aggregierte Daten, welches auf Berry (1994) und Berry *et al.* (1995) zurückgeht und bspw. ausführlich von Nevo (2000) diskutiert wird. Dieses Ausgangsmodell wird um *zeitvariable Parameter* erwei-

tert. Das Preismodell erfüllt den Zweck, das (u. U. vorliegende) Endogenitätsproblem im Nachfragemodell mithilfe des CF-Ansatzes zu beheben (Kim 2010).

3.2.1 Nachfragemodell

Angenommen, es gibt $i = 1, \dots, I$ Konsumenten, die aus $j = 1, \dots, J$ Marken zu den Zeitpunkten $t = 1, \dots, T$ genau die Marke auswählen, die ihnen den höchsten Nutzen u_{ijt} stiftet. Um Kategoriekäufe abzubilden, wird zusätzlich mit $j = 0$ die *Nichtwahloption* modelliert (Chintagunta 2001). Ausgangspunkt für die nächsten Abschnitte ist die (indirekte) Nutzenfunktion:³

$$u_{ijt} = \begin{cases} \mathbf{x}'_{jt} \boldsymbol{\alpha}_{it} + \beta_{it} \ln(p_{jt}) + \varepsilon_{jt} + \nu_{ijt}, & j = 1, \dots, J \\ \nu_{ijt}, & j = 0, \end{cases} \quad (3.1)$$

mit $i = 1, \dots, I$ und $t = 1, \dots, T$.

Bei \mathbf{x}_{jt} handelt es sich um einen $K \times 1$ -Vektor mit exogenen Variablen (z. B. Promotionaktivitäten) und Indikatorvariablen für die Markenkonstanten.⁴ Die logarithmierte Preisvariable $\ln(p_{jt})$ ist potenziell *endogen* und wird aus didaktischen Gründen gesondert betrachtet. Insgesamt gibt es also $K + 1$ Parameter in der Nutzenfunktion. Des Weiteren existieren mit ε_{jt} und ν_{ijt} zwei Störterme, die unterschiedliche Phänomene abbilden. Über $\varepsilon_{jt} \sim N(0, \sigma_{\varepsilon_j}^2)$ werden (für den Analytiker) *unbeobachtbare Markencharakteristika (UMC)* modelliert. Es erscheint unwahrscheinlich, dass stets alle relevanten Größen, die die Nachfrage beeinflussen, im Modell enthalten sind. Werden aber solche *Nachfrageschocks* vernachlässigt, können verzerrte Ergebnisse resultieren (Chintagunta *et al.* 2005; Park und Gupta 2009). Hinzu kommt, dass Nachfrageschocks aufgrund von UMC höchstwahrscheinlich zu einem Endogenitätsproblem führen und dann i. d. R. $E[\varepsilon_{jt} \ln(p_{jt})] \neq 0$ gilt (Petrin und Train 2010). Folglich ist der Preiseffekt *verzerrt*. Der Störterm ν_{ijt} ist unabhängig identisch extremwertverteilt und über ihn werden personen-, marken- und zeitspezifische Nutzenabweichungen abgebildet. Durch ε_{jt} und ν_{ijt} hält Zufälligkeit aus Sicht des

³ Im vorliegenden Beitrag kommen *logarithmierte* Preise zum Einsatz, weil diese in der empirischen Studie einen *besseren Datenfit* liefern. Modellherleitung und -beschreibung gelten aber gleichermaßen für Nutzenfunktionen mit linearem Preiseffekt.

⁴ Im vorliegenden Beitrag wird auch dann der Terminus *Markenkonstante* verwendet, wenn diese *zeitvariabel* ist.

Analytikers Einzug in die Nutzenfunktion. Der Entscheider ist sich „seiner“ Werte für ε_{jt} und v_{ijt} voll bewusst.

Die zu schätzenden Parameter der Nutzenfunktion (3.1) sind *personen- und zeitspezifisch* und besitzen folglich die Indizes i und t . Die Heterogenität wird kontinuierlich über die *Normalverteilung* modelliert. Die Annahme der Normalverteilung ist im Kontext des vorliegenden Beitrags sehr praktisch, weil hierdurch Dynamik (d.h. Änderungen der Mittelwerte) und Heterogenität (d.h. Abweichungen einzelner Werte vom Mittelwert) *additiv separabel* sind.

Die Parameter sind für $i = 1, \dots, I$ und $t = 1, \dots, T$ wie folgt verteilt:

$$\alpha_{it} = \bar{\alpha}_t + \mathbf{v}_{\alpha i} \text{ und} \quad (3.2)$$

$$\beta_{it} = \bar{\beta}_t + v_{\beta i}, \text{ mit } [\mathbf{v}_{\alpha i} \ v_{\beta i}]' = \mathbf{v}_i \sim N(\mathbf{0}, \mathbf{\Psi}). \quad (3.3)$$

Die Mittelwerte $\bar{\alpha}_t$ und $\bar{\beta}_t$ folgen einem *Random Walk-Prozess*:⁵

$$\bar{\alpha}_t = \bar{\alpha}_{t-1} + \boldsymbol{\eta}_{\alpha t} \text{ und} \quad (3.4)$$

$$\bar{\beta}_t = \bar{\beta}_{t-1} + \eta_{\beta t}, \text{ mit } [\boldsymbol{\eta}_{\alpha t} \ \eta_{\beta t}]' = \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}). \quad (3.5)$$

Im Gegensatz zu einem Modell mit konstanten Parametern gibt es also Variation sowohl in der Zeitdimension (*unbeobachtbare Dynamik*) als auch in der Personendimension (*unbeobachtbare Heterogenität*). Die gewählte Spezifikation ist äußerst *flexibel* und erlaubt zudem, im Rahmen *genesteter Modelle* unterschiedliche Spezifikationen (mit/ohne Dynamik und/oder Heterogenität) gegeneinander zu testen.

Für ein besseres Verständnis des Modells und vor allem des nachfolgend beschriebenen Vorgehens zur Modellschätzung ist es praktikabel und in der Literatur üblich, die Komponenten der Nutzenfunktion danach zu strukturieren, ob sie konsumentenspezifisch sind oder nicht (Nevo 2000). Einsetzen von (3.2) und (3.3) in die Nutzenfunktion (3.1) führt zu:

⁵ Es sind alternative Spezifikationen für die Parameterdynamiken möglich, z. B. in Abhängigkeit von beobachteten exogenen Einflussgrößen oder in Form autoregressiver Prozesse höherer Ordnung (vgl. Abschnitt 2.2 und Lachaab *et al.* 2006). Für die Fragestellungen des vorliegenden Beitrags sind diese Spezifikationen aber nicht relevant und werden daher nicht behandelt.

$$u_{ijt} = \mathbf{x}'_{jt}(\bar{\boldsymbol{\alpha}}_t + \mathbf{v}_{\alpha i}) + (\bar{\beta}_t + v_{\beta i}) \ln(p_{jt}) + \varepsilon_{jt} + v_{ijt}, \quad (3.6)$$

$$\Leftrightarrow u_{ijt} = \underbrace{\mathbf{x}'_{jt}\bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t \ln(p_{jt}) + \varepsilon_{jt}}_{\delta_{jt}} + \underbrace{\mathbf{x}'_{jt}\mathbf{v}_{\alpha i} + v_{\beta i} \ln(p_{jt})}_{\mu_{ijt}} + v_{ijt}, \quad (3.7)$$

$$\Leftrightarrow u_{ijt} = \delta_{jt} + \mu_{ijt} + v_{ijt}, \quad (3.8)$$

mit $i = 1, \dots, I$, $j = 1, \dots, J$ und $t = 1, \dots, T$.

Den *Durchschnittsnutzen aller Personen* fasst δ_{jt} zusammen. Dagegen beinhalten $\mu_{ijt} + v_{ijt}$ alle *personenspezifischen* (heteroskedastischen) Abweichungen von δ_{jt} , die der Analytiker *nicht* beobachtet. Es ist wichtig zu verstehen, dass δ_{jt} sämtliche Variation in den Dimensionen j (Marken) und t (Perioden) aufnimmt. Bedingt auf δ_{jt} spielen Verzerrungen aufgrund von UMC, Endogenität und Dynamik in der Nutzenfunktion folglich keine Rolle.

Die angenommene *Extremwertverteilung* von v_{ijt} erlaubt indes die geschlossene Berechnung der Wahlwahrscheinlichkeit pr_{ijt} , dass der *nutzenmaximierend* agierende Konsument i die Marke j zum Zeitpunkt t wählt. Es folgt das bekannte *Logitmodell* (vgl. z. B. ausführlich Train 2009):

$$\text{pr}_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k=1}^J \exp(\delta_{kt} + \mu_{ikt})}, \quad (3.9)$$

mit $i = 1, \dots, I$, $j = 1, \dots, J$ und $t = 1, \dots, T$.

Aggregierte Marktanteile s_{jt} resultieren aus der *Integration* über die *unbeobachteten* heterogenen Abweichungen der Konsumenten. Da die Heterogenität in μ_{ijt} aus \mathbf{v}_i resultiert, gilt:

$$s_{jt} = \int_{-\infty}^{\infty} \text{pr}_{ijt}(\mathbf{v}_i) \phi(\mathbf{v}_i | \mathbf{0}, \boldsymbol{\Psi}) d\mathbf{v}_i. \quad (3.10)$$

Dabei bezeichnet ϕ die Dichtefunktion einer (multivariaten) Normalverteilung. Für das Integral dieser Marktanteilsleichung gibt es keine geschlossene Lösung, daher sind *Monte Carlo-Techniken* notwendig, um die Marktanteile zu bestimmen. Wenn \mathbf{v}_s die s -te Ziehung von \mathbf{v} aus der Verteilung $N(\mathbf{0}, \boldsymbol{\Psi})$ bezeichnet und insg. S Zufallsziehungen verwendet werden, dann gilt für die Marktanteile:

$$s_{jt} \approx \frac{1}{S} \sum_{s=1}^S \frac{\exp(\delta_{jt} + \mu_{sjt}(\mathbf{v}_s))}{1 + \sum_{k=1}^J \exp(\delta_{kt} + \mu_{skt}(\mathbf{v}_s))}, \quad (3.11)$$

mit $j = 1, \dots, J$ und $t = 1, \dots, T$.

Im vorliegenden Beitrag werden Zufallszahlen basierend auf *Haltonsequenzen* eingesetzt, die günstige Eigenschaften für die Modellierung und Schätzung von Heterogenität aufweisen und mit einer relativ geringen Anzahl von Ziehungen präzisere Ergebnisse für die Berechnung der Marktanteilsintegrale erreichen, als „normale“ Pseudo-Zufallszahlen (Train 2009).

Die Nachfrage ergibt sich aus der Multiplikation des Marktanteils gemäß (3.11) mit dem Marktvolumen zum Zeitpunkt t :

$$q_{jt} = M_t \cdot s_{jt}. \quad (3.12)$$

3.2.2 Preismodell

Bei der Modellierung der Preisfunktion stellt sich die Frage, wodurch die Variation in den beobachteten Preisen zustande kommt. In der Literatur werden hierzu zwei verschiedene Ansätze unterschieden: *Strukturelle* und *deskriptive* Preismodelle. Bei strukturellen Preismodellen wird angenommen, dass die Preissetzung der Anbieter auf ein *Optimierungskalkül* gründet, z. B. der Gewinnmaximierung von Herstellern und Händlern (Yang *et al.* 2003; Draganska und Klapper 2007). Dieses Vorgehen besitzt die *Vorteile*, dass

1. neben der Nachfrage *auch* die Preise *strukturell* modelliert werden,
2. sofern Nachfrage- und Preisfunktion gleichzeitig geschätzt werden, *statistische Effizienzvorteile* in der Schätzung erzielt werden können und
3. sich *Preisendogenitätsprobleme* damit *lösen lassen*, weil explizit modelliert wird, warum $\mathbb{E}[\varepsilon_{jt} | \ln(p_{jt})] \neq 0$ gilt.

Ein Nachteil ist allerdings, dass *sämtliche* Ergebnisse *verzerrt* sein können, sofern die getroffenen Annahmen bzgl. des Optimierungskalküls falsch sind. Dieser Punkt wird ausführlich von Dubé und Chintagunta (2003,

S. 294 f.) im Kommentar zu Yang *et al.* (2003) diskutiert. Des Weiteren ist die Schätzung, insb. mit likelihoodbasierten Verfahren und in Verbindung mit dem Nachfragemodell, sehr aufwendig und kompliziert (Yang *et al.* 2003, S. 258 f.).

Deskriptive („reduced form“) Preismodelle sind dagegen weniger restriktiv, weil sie auf die Annahme eines bestimmten Optimierungskalküls verzichten. Dieses Vorgehen wird im vorliegenden Beitrag präferiert, weshalb das angesprochene Problem möglicher Verzerrungen aufgrund eines fehlspezifizierten Preissetzungsmechanismus nicht besteht. Trotzdem ist das Endogenitätsproblem auf diese Weise zu beheben und es gilt (Train 2009):

$$\ln(p_{jt}) = \mathbf{z}_{jt}'\boldsymbol{\gamma} + \zeta_{jt}, \text{ mit } j = 1, \dots, J \text{ und } t = 1, \dots, T. \quad (3.13)$$

Der (logarithmierte) Preis wird demnach als lineare Funktion der *exogenen Instrumentalvariablen* (\mathbf{z}_{jt}) spezifiziert, wobei es pro endogener Variable wenigstens eine Instrumentalvariable geben muss. In \mathbf{z}_{jt} sind *zudem* alle exogenen Variablen des Nachfragemodells (\mathbf{x}_{jt}) enthalten sowie ggf. nicht-lineare Transformationen und Interaktionen der Variablen in \mathbf{z}_{jt} , um möglichst viel *exogene* Variation in $\ln(p_{jt})$ zu erklären (Jiang *et al.* 2009). Welche Variablen als Instrumente dienen, hängt von der jeweiligen Anwendung ab.⁶ Typische Instrumente sind *Kostengrößen* (Chintagunta 2001), *zeitverzögerte Preise* (Villas-Boas und Winer 1999) oder *Preise anderer (geografischer) Märkte* (Nevo 2000). Bei allen diesen Variablen wird davon ausgegangen, dass sie zwar mit dem Preis, aber *nicht* mit dem Störterm ε_{jt} korrelieren.

Mögliche perioden- und markenspezifische Abweichungen der Preisvariationen, von denen der exogenen Instrumente, werden durch den Störterm ζ_{jt} abgefangen, für den $\zeta_{jt} \sim N(0, \sigma_\zeta^2)$ gilt. Sofern alle Instrumente die *Exogenitätsbedingung* erfüllen, sind mögliche Korrelationen zwischen den UMC und den Preisen auf Preisvariationen zurückzuführen, welche ζ_{jt} verursacht. Die Grundidee von (3.13) ist folglich die Trennung der gesamten Preisvariation in den Teil, der durch exogene Variablen erklärbar ist und den *unverfälschten* Preiseffekt ($\mathbf{z}_{jt}'\boldsymbol{\gamma}$) darstellt sowie den Teil, der das Endogenitätsproblem verursacht (ζ_{jt}).

⁶ Vgl. hierzu Abschnitt 3.5.2 des vorliegenden Beitrags.

3.2.3 Kontrollfunktionenansatz

Wie zuvor erwähnt, liegt ein Preisendogenitätsproblem genau dann vor, wenn $\mathbb{E}[\varepsilon_{jt} | \ln(p_{jt})] \neq 0$ gilt. Mögliche Gründe für die hieraus resultierenden Korrelationen zwischen den UMC und den Preisen sind vielfältig (Train 2009, S. 315 f.). Beispielsweise ist es denkbar, dass eine eher hochpreisige Marke einen deutlich besseren Regalplatz im Handel bekommt als ein Billigprodukt. Hieraus resultiert eine *positive* Korrelation zwischen ε_{jt} und $\ln(p_{jt})$ und der gemessene Preiseffekt ist ohne Endogenitätsbehandlung zu *niedrig*. Des Weiteren kann es sein, dass (nicht beobachtete) Promotionmaßnahmen, z. B. Sonderplatzierungen, mit Preissenkungen einhergehen. Folglich liegt hier eine *negative* Korrelation zwischen ε_{jt} und $\ln(p_{jt})$ vor und Preiseffekte würden *überschätzt*. Es sei an dieser Stelle hervorgehoben, dass das Endogenitätsproblem nicht gelöst werden kann, indem der UMC-Term (wie bis vor wenigen Jahren in der Marketingliteratur üblich) vernachlässigt wird. Zumindest ein Teil der Variation der UMC wird dann vom Logitstörterm ν_{ijt} aufgefangen und damit ist das Problem noch immer gegenwärtig (Park und Gupta 2009; Petrin und Train 2010). Das Endogenitätsproblem lässt sich auch nicht lösen, indem größere Datenmengen (= mehr Beobachtungen) verwendet werden, da hierdurch die erwähnte Korrelation zwischen ε_{jt} und $\ln(p_{jt})$ *nicht abnimmt* oder gar *verschwindet*.

Eine einfache, aber sehr flexible Möglichkeit, den Zusammenhang zwischen den UMC und den Preisen abzubilden, ist die Annahme, dass die beiden Störterme von Preis- und Nachfragefunktion korreliert sind (Villas-Boas und Winer 1999). Wegen der Verteilungsannahmen für ζ_{jt} und ε_{jt} folgt eine *multivariate Normalverteilung* (Kim und Kim 2011):

$$\begin{bmatrix} \zeta_{jt} \\ \varepsilon_{jt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\zeta_j}^2 & \rho_j \sigma_{\zeta_j} \sigma_{\varepsilon_j} \\ \rho_j \sigma_{\varepsilon_j} \sigma_{\zeta_j} & \sigma_{\varepsilon_j}^2 \end{bmatrix} \right), \quad (3.14)$$

für $j = 1, \dots, J$ und $t = 1, \dots, T$.

Die Kovarianzmatrix in 3.14 sei mit $\mathbf{\Omega}$ bezeichnet und sämtliche Parameter sind markenspezifisch. Vor allem für die Korrelationen ρ_j ist das wichtig, denn auf diese Weise wird erlaubt, dass einzelne Marken ein Endogenitätsproblem besitzen können ($\rho_j \neq 0$) und andere nicht ($\rho_j = 0$) (Park und Gupta 2009). Eine *Cholesky-Zerlegung* ($\mathbf{\Omega} = \mathbf{U}' \cdot \mathbf{U}$) ermöglicht, die beiden korrelierten Störterme, ζ_{jt} und ε_{jt} , in die *unkorrelierten* und jeweils *standardnormalverteilten* Störterme, ζ_{jt}^* und ε_{jt}^* , zu *dekomponieren* (Train

2009, S. 208 f.). Es gilt:

$$\begin{bmatrix} \zeta_{jt} \\ \varepsilon_{jt} \end{bmatrix} = \begin{bmatrix} \sigma_{\zeta j} & 0 \\ \rho_j \sigma_{\varepsilon j} & \sqrt{1 - \rho_j^2} \sigma_{\varepsilon j} \end{bmatrix} \cdot \begin{bmatrix} \zeta_{jt}^* \\ \varepsilon_{jt}^* \end{bmatrix}, \text{ mit } \begin{bmatrix} \zeta_{jt}^* \\ \varepsilon_{jt}^* \end{bmatrix} \sim N(\mathbf{0}, \mathbf{I}_2), \quad (3.15)$$

für $j = 1, \dots, J$ und $t = 1, \dots, T$.

Hierdurch lässt sich die Nutzenfunktion derart aufstellen, dass die Bedingung $E[\varepsilon_{jt}^* | \ln(p_{jt})] = 0$ stets erfüllt ist. Weil – wie bereits erwähnt – die gesamte Endogenitätsproblematik durch *Durchschnittsnutzen* aufgefangen wird, beschränken sich die weiteren Erläuterungen auf δ_{jt} .

$$\delta_{jt} = \mathbf{x}_{jt}' \bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t \ln(p_{jt}) + \rho_j \sigma_{\varepsilon j} \zeta_{jt}^* + \sqrt{1 - \rho_j^2} \sigma_{\varepsilon j} \varepsilon_{jt}^*, \quad (3.16)$$

mit $j = 1, \dots, J$ und $t = 1, \dots, T$.

Der mit dem Preis korrelierte Teil in ε_{jt} wird folglich *herausgetrennt*. Der Preis ist zwar noch immer mit ζ_{jt}^* korreliert, hierfür gibt es aber eine Schätzung, die bspw. durch eine OLS-Regression von $\ln(p_{jt})$ auf \mathbf{z}_{jt} gebildet wird. Aus Gleichung (3.16) ist direkt ersichtlich, dass für $\rho_j = 0$ (keine Korrelation der Störterme) und/oder $\sigma_{\varepsilon j} = 0$ (keine UMC) kein Endogenitätsproblem existiert. Ausgehend von der Preisgleichung (3.13) lässt sich ζ_{jt}^* berechnen:

$$\zeta_{jt}^* = \sigma_{\zeta j}^{-1} \left(\ln(p_{jt}) - \mathbf{z}_{jt}' \boldsymbol{\gamma} \right). \quad (3.17)$$

Dieser Residualterm wird als *Kontrollfunktion* (CF) bezeichnet (Heckman 1978; Hausman 1978; Petrin und Train 2010) und Einsetzen in (3.16) führt zu:

$$\delta_{jt} = \mathbf{x}_{jt}' \bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t \ln(p_{jt}) + \rho_j \sigma_{\varepsilon j} \sigma_{\zeta j}^{-1} \left(\ln(p_{jt}) - \mathbf{z}_{jt}' \boldsymbol{\gamma} \right) + \sqrt{1 - \rho_j^2} \sigma_{\varepsilon j} \varepsilon_{jt}^*, \quad (3.18)$$

mit $j = 1, \dots, J$ und $t = 1, \dots, T$.

Die Idee des CF-Ansatzes ist also, dass – bedingt auf die Residuen der Preisgleichung – der ursprüngliche Preis *unkorreliert* mit dem Störterm der Nutzenfunktion ε_{jt}^* ist. Das Endogenitätsproblem mithilfe von CF zu

lösen, hat für den vorliegenden Beitrag *zwei entscheidende Vorteile*: Der CF-Ansatz ist auch in Modellen mit *zeitvariablen Parametern* einsetzbar und bietet eine einfache Möglichkeit für einen *Endogenitätstest*. Beide Vorteile seien nachfolgend kurz erläutert.⁷

Das Problem der Standardverfahren

Zur Lösung des Endogenitätsproblems auf Nachfrageebene erfolgt i. d. R. eine *Bereinigung* der betroffenen Variable(n). Hierzu kommen *Standardverfahren* wie die zweistufige kleinste Quadrate-Schätzung (2SLS) oder die generalisierte Momentenmethode (GMM) zum Einsatz (Greene 2011, S. 259 ff. und S. 495 ff.). Diese Verfahren sind allerdings nur dann einsetzbar, wenn zwei Bedingungen erfüllt sind:

1. Das *Modell* muss *linear* (in den Parametern) sein.
2. Der *Parameter der endogenen Variable* muss *zeitlich konstant* sein.

Da das Modell von Berry *et al.* (1995) beide Bedingungen erfüllt, sind die genannten Standardverfahren zulässig. In der Tat ist ein Hauptargument für das im Abschnitt (3.3) skizzierte Vorgehen zur Schätzung diskreter Wahlmodelle für aggregierte Daten durch eine Linearisierung des nicht-linearen Logitmodells, die Möglichkeit zur Berücksichtigung der Preisendogenität mit Standardverfahren (Berry 1994).

Die Bedingung der Linearität alleine reicht allerdings nur dann, wenn der Effekt der endogenen Variablen (zeitlich) konstant ist. Eine weitere und nur für Modelle mit zeitvariablen Parametern relevante Bedingung, die erfüllt sein muss, damit es zu keinen Endogenitätsverzerrungen kommt, lautet: $\mathbb{E}[\varepsilon_t | \bar{\beta}_t] = 0$. Es gilt also nicht nur die *Exogenitätsbedingung* für die Variablen des Modells, sondern ebenfalls für die Parameter.

Zur Verdeutlichung sei exemplarisch die Invalidität des 2SLS-Schätzers im Falle zeitvariabler Parameter anhand eines Beispiels beschrieben. Hierzu ist anstelle der CF der *prognostizierte Preis* $\ln(p_{jt}) = \mathbf{z}'_{jt}\boldsymbol{\gamma}$ in (3.16) einzusetzen und es folgt:

$$\delta_{jt} = \mathbf{x}'_{jt}\boldsymbol{\alpha}_t + \bar{\beta}_t \ln(p_{jt}) + \varepsilon_{jt}^\dagger \text{ und} \quad (3.19)$$

$$\varepsilon_{jt}^\dagger = \varepsilon_{jt} + \bar{\beta}_t (\ln(p_{jt}) - \ln(p_{jt})), \text{ mit} \quad (3.20)$$

$$j = 1, \dots, J \text{ und } t = 1, \dots, T.$$

⁷ Die Ausführungen basieren vorrangig auf Kim (2010) sowie Kim und Kim (2011).

Bedingt auf den *endogenitätsbereinigten Preis* ist $\tilde{\beta}_t$ mit dem Störterm ε_{jt}^\dagger korreliert und erzeugt ein *weiteres Endogenitätsproblem*. Eine Betrachtung von (3.18) deckt auf, dass der beschriebene CF-Ansatz dieses Problem *nicht besitzt*. Weder $\tilde{\beta}_t$ noch $\ln(p_{jt})$ sind in (3.18) mit dem Störterm korreliert und daher handelt es sich beim CF-Ansatz um ein valides Verfahren, um dem Endogenitätsproblem in Logitmodellen mit zeitvariablen Parametern zu begegnen.

Endogenitätstest

Der vorgestellte CF-Ansatz erlaubt einen Test auf Endogenität direkt über die geschätzten Parameter ρ_j . Unter der Nullhypothese eines nicht vorhandenen Endogenitätsproblems gilt $H_0: \rho_j = 0$. Dies lässt sich pro Marke mit einem t -Test prüfen. Für alle Marken gleichzeitig kommt bei einer Maximum-Likelihood-Schätzung z. B. der *Likelihood-Quotienten-Test* in Frage (Greene 2011, S. 566 f.). Solche Endogenitätstests können als Erweiterung des Hausmann-Tests (Hausman 1978) verstanden werden, für den Fall von aggregierten Logitmodellen mit Heterogenität *und* zeitvariablen Parametern. Der „normale“ Hausman-Test (Greene 2011, S. 274 f.), der i. d. R. als Endogenitätstest Anwendung findet, ist für Modelle mit zeitvariablen Parametern nicht anwendbar, da er explizit auf den geschätzten Parametern der Modelle beruht und es hiervon entsprechend T Stück gibt.

3.2.4 Zustandsraummodell

Zur Schätzung der zeitvariablen Parameter wird das zuvor hergeleitete Nachfragemodell (inkl. CF) kompakt als ZRM formuliert (Harvey 1989). Der Vorteil liegt darin, dass sich auf ZRM direkt die KF-Rekursion zur Ermittlung der zeitvariablen Parameter anwenden lässt (Kalman 1960). ZRM bestehen aus zwei gekoppelten (linearen) Matrixgleichungen (siehe Abschnitt 2.2): Der *Beobachtungsgleichung*, über die ein dynamischer Zusammenhang „gemessen“ wird, und der *Zustandsgleichung*, die den dynamischen Zusammenhang selbst abbildet. Für das Nachfragemodell lautet das ZRM $\forall t$:

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{a}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad \text{mit } \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{R}_t) \quad \text{und} \quad (3.21)$$

$$\mathbf{a}_t = \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \quad \text{mit } \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t). \quad (3.22)$$

Im Kontext des vorliegenden Beitrags bilden die J Durchschnittsnutzenfunktionen (3.16 bzw. 3.18) die Beobachtungsgleichung (3.21), weshalb

$\mathbf{y}_t = [\delta_{1t} \dots \delta_{Jt}]'$ gilt. Die Matrix \mathbf{H}_t besitzt die Dimension $J \times (K + 1)$ und beinhaltet alle Einflussgrößen mit einem zeitvariablen Effekt.⁸ Hierzu gehören die Indikatorvariablen für die Markenkonstanten sowie die Marketing-Mix-Variablen. Im $J \times 1$ -Vektor \mathbf{d}_t sind exogene Einflüsse auf den Durchschnittsnutzen zu finden, die *keinen* zeitvariablen Effekt besitzen. Das wären hier die J Effekte der CF: $\mathbf{d}_t = [\rho_1 \sigma_{\varepsilon 1} \zeta_{1t}^* \dots \rho_J \sigma_{\varepsilon J} \zeta_{Jt}^*]'$. Entsprechend gilt für den $J \times 1$ -Vektor des Störterms $\boldsymbol{\varepsilon}_t = [\varepsilon_{1t} \dots \varepsilon_{Jt}]'$ und dessen Kovarianzmatrix $\mathbf{R}_t = \mathbf{I}_J \cdot [(1 - \rho_1^2) \sigma_{\varepsilon 1}^2 \dots (1 - \rho_J^2) \sigma_{\varepsilon J}^2]'$.

Die Zustandsgleichung (3.22) bildet die Dynamik in den Parametern gemäß (3.4) bzw. (3.5) ab. Daher sind alle zeitvariablen Parameter im $(K + 1) \times 1$ -Vektor $\mathbf{a}_t = [\bar{\boldsymbol{\alpha}}_t \ \bar{\boldsymbol{\beta}}_t]'$ zusammengefasst. Für den Störterm $\boldsymbol{\eta}_t$ und die dazugehörige Kovarianzmatrix \mathbf{Q}_t gelten die Annahmen gemäß (3.5).

3.3 Schätzung

Dieser Abschnitt beschreibt das Vorgehen zur *Schätzung* des zuvor präsentierten Modells. Die Schätzung wird *frequentistisch* durchgeführt und es sind *zwei Gruppen von Parametern* zu unterscheiden: Zum einen gibt es *zeitvariable Parameter* (im Nachfragemodell $\bar{\boldsymbol{\alpha}}_t$ und $\bar{\boldsymbol{\beta}}_t$), zum anderen sind auch *konstante Parameter* vorhanden (im Nachfragemodell ρ_j und $\sigma_{\varepsilon j}$ sowie die Varianzparameter in $\boldsymbol{\Omega}$ und $\boldsymbol{\Psi}$ und im Preismodell $\sigma_{\zeta j}$ und $\boldsymbol{\gamma}$).⁹

Zur Schätzung der konstanten Parameter wird ein *likelihoodbasierter Ansatz* verfolgt und die Konstruktion der Likelihoodfunktion für das gesamte Modell wird in Abschnitt 3.3.1 erläutert. Bedingt auf Werte für die geschätzten konstanten Parameter, lassen sich anschließend die zeitvariablen Parameter mithilfe des KF¹⁰ bzw. *Kalman Smoothers* (KS) ermitteln, was Abschnitt 3.3.2 zeigt.

⁸ Vgl. zur Überführung des Logitmodells für aggregierte Daten mit zeitvariablen Parametern in die Form eines ZRM auch Abschnitt (2.5) der vorliegenden Arbeit.

⁹ Eine Erweiterung mit zeitvariablen Parametern im Preismodell wäre ebenfalls möglich (Kim 2010).

¹⁰ Vgl. hierzu auch ausführlich Abschnitt 2.3.

3.3.1 Konstante Parameter

Sämtliche konstanten Parameter im Nachfragemodell ($\boldsymbol{\rho}$, $\boldsymbol{\sigma}_\varepsilon$, $\text{vec}(\boldsymbol{\Psi})$, $\bar{\boldsymbol{\alpha}}_0$, $\bar{\boldsymbol{\beta}}_0$, $\text{vec}(\mathbf{Q})$) und Preismodell ($\boldsymbol{\gamma}$, $\boldsymbol{\sigma}_\zeta$) werden *gleichzeitig* geschätzt, um das *Problem generierter Regressoren* zu vermeiden (Pagan 1984). Würden die CF zuerst separat ermittelt und anschließend in das Nachfragemodell als Regressoren eingesetzt, wären die Standardfehler der zweiten Stufe falsch aufgrund einer Vernachlässigung der Schätzunsicherheit. Zudem kommt es in dynamischen Modellen ebenfalls zu einer Verzerrung der zeitvariablen Parameter (Kim 2010; Kim und Kim 2011).

Weil zur Berechnung der Marktanteile in (3.10) Simulationstechniken notwendig sind, handelt es sich de facto um einen *Maximum-Simulated-Likelihood (MSL)-Schätzer* (Train 2009, S. 237 ff.). Hierbei werden die Verteilungsannahmen bzgl. der Störterme des Nachfrage- und Preismodells sowie der zeitvariablen Parameter genutzt, um die gemeinsame Verteilung der unbekannten Parameter gegeben der Daten zu bestimmen. Zur Herleitung der simulierten (Log-)Likelihoodfunktion sind *vier Schritte* notwendig: die *Marktanteilsinvertierung*, die *Regression* für das *Preismodell*, die *Kalman Filter-Schätzung* für das *Nachfragemodell* und die Berechnung der *Jacobi-Determinante*. Diese vier Schritte, sowie deren Verbindung untereinander sind synoptisch Abbildung 3.1 zu entnehmen und werden nachfolgend detailliert diskutiert.

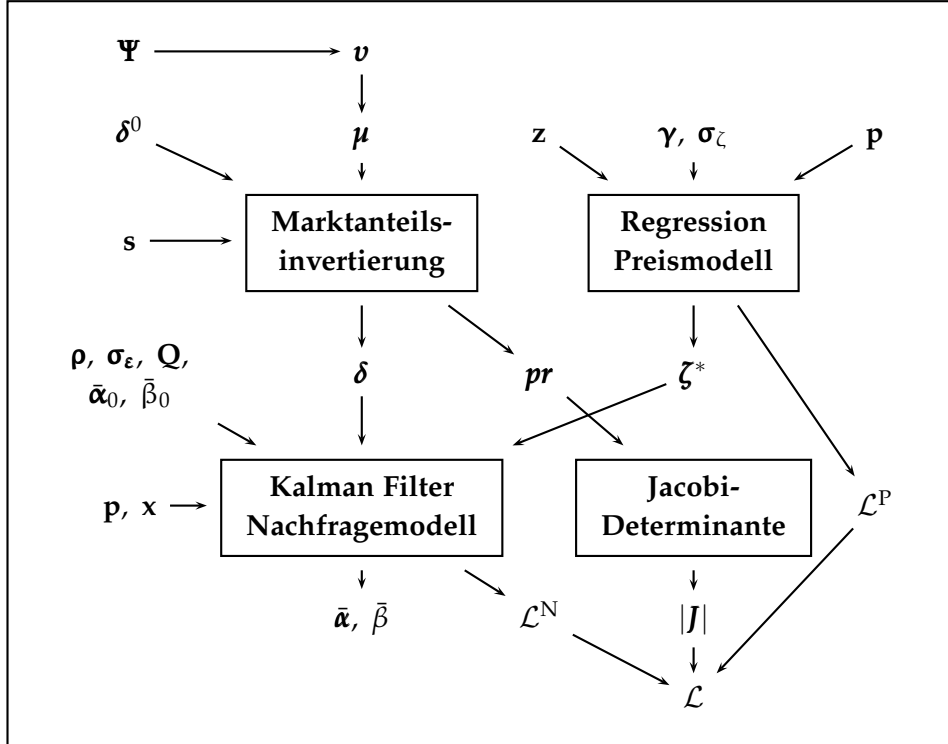
Marktanteilsinvertierung

Im ersten Schritt erfolgt die *Marktanteilsinvertierung*, um δ_{jt} zu erhalten (Berry 1994).¹¹ Diese Durchschnittsnutzen dienen dann dem Nachfragemodell als abhängige Variable mit den UMC als Störterm. Für gegebene Heterogenitätsparameter in $\boldsymbol{\Psi}$ und Ziehungen der Zufallszahlen lassen sich alle μ_{sjt} ermitteln. Berry *et al.* (1995, S. 863 ff.) schlagen nun vor, δ_{jt} mithilfe einer *kontrahierenden Abbildung* zu bestimmen. Die Autoren beweisen, dass δ_{jt} *eindeutig* ermittelt werden kann und zwar derart, dass geschätzte und beobachtete Marktanteile exakt einander entsprechen. Die Berechnung erfolgt *iterativ* über die Gleichung (vgl. Train 2009, S. 322):

$$\delta_{jt}^{(h+1)} = \delta_{jt}^{(h)} + \ln \left(\frac{s_{jt}}{s_{jt}(\boldsymbol{\delta}_t^{(h)})} \right), \text{ mit } j = 1, \dots, J \text{ und } t = 1, \dots, T. \quad (3.23)$$

¹¹ Vgl. hierzu ebenfalls Abschnitt 2.5 im 2. Kapitel der vorliegenden Arbeit.

Abbildung 3.1: Konzeption der Parameterschätzung



Anmerkung: Die einzelnen Größen, z. B. $\boldsymbol{\mu}$ oder $\boldsymbol{\delta}$, beinhalten immer sämtliche relevanten Werte für $j = 1, \dots, J$ und $t = 1, \dots, T$ (sowie ggf. für $s = 1, \dots, S$). Dies gilt auch für die Inputdaten (\mathbf{s} , \mathbf{p} , \mathbf{x} und \mathbf{z}). Alle dargestellten Schritte müssen für jede Iteration in der Likelihoodmaximierung zur Schätzung der konstanten Parameter ($\boldsymbol{\rho}$, σ_ϵ , $\text{vec}(\boldsymbol{\Psi})$, $\bar{\alpha}_0$, $\bar{\beta}_0$, $\text{vec}(\mathbf{Q})$, $\boldsymbol{\gamma}$, σ_ζ) durchlaufen werden.

Bei s_{jt} und \hat{s}_{jt} handelt es sich um die wahren bzw. geschätzten Marktanteilswerte und letztere sind eine Funktion der geschätzten Durchschnittsnutzen sowie der heterogenen Abweichungen (vgl. Gleichung 3.11). Gleichung (3.23) wird so lange ausgeführt, bis die Differenz zwischen $\delta_{jt}^{(h+1)}$ und $\delta_{jt}^{(h)}$ hinreichend klein ist.¹² Sofern keine Heterogenität vorliegt, ist δ_{jt} ohne Schätzung über eine simple Transformation der Marktanteilsdaten zu erhalten. Dann gilt der Zusammenhang $\delta_{jt} = \ln(s_{jt}/s_{0t})$.¹³ Diese homogenen Durchschnittsnutzen sind für heterogene Modelle auch als Startwerte für (3.23) nützlich: $\delta_{jt}^{(0)} = \delta_{jt}$.

¹² In Anlehnung an die Empfehlung von Dubé *et al.* (2012), wird ein sehr strenges Kriterium gewählt: $\|\delta_{jt}^{(h+1)} - \delta_{jt}^{(h)}\| < 1 \cdot 10^{-14}$, $\forall j, t$.

¹³ Diese „Invertierung der Marktanteile“ ist durch Berry (1994) sehr populär geworden. Gleichwohl existiert mit dem *Log-Centering-Ansatz* von Nakanishi und Cooper (1974) ein ähnliches Vorgehen in der Marketingliteratur zur Linearisierung von Marktanteilsmodellen schon seit den 1970er Jahren.

Regression für das Preismodell

Um eine Schätzung für die CF (ζ_{jt}) zu erhalten, wird $\ln(p_{jt})$ auf z_{jt} regressiert und das Ergebnis gemäß (3.17) verwendet. Da das Preismodell (3.13) linear in ζ_{jt}^* ist, folgt die Likelihoodfunktion sofort aus der Annahme der Normalverteilung des Störterms (Greene 2011, S. 550):

$$\mathcal{L}^P = \ln L(\boldsymbol{\theta}^P | \mathbf{p}, \mathbf{z}) = \ln \left(\prod_{t=1}^T \prod_{j=1}^J f(p_{jt}, z_{jt} | \boldsymbol{\theta}^P) \right) \quad (3.24)$$

$$= -0,5 \cdot \sum_{t=1}^T \sum_{j=1}^J \left[\ln(2\pi\sigma_{\zeta_j}^2) + \zeta_{jt}^2 / \sigma_{\zeta_j}^2 \right]. \quad (3.25)$$

Der Vektor $\boldsymbol{\theta}^P = [\boldsymbol{\gamma} \ \boldsymbol{\sigma}_{\zeta}]'$ enthält alle zu schätzenden konstanten Parameter des Preismodells. Die Reihenfolge der Summen in (3.25) ist irrelevant, weil das Preismodell weder perioden- noch markenübergreifende Effekte enthält.

Kalman Filter-Schätzung für das Nachfragemodell

Es bietet sich an, die Likelihoodfunktion des Nachfragemodells mithilfe der sog. *Prognosefehler-Zerlegung* aufzustellen (Schweppe 1965). Die Idee ist hierbei, dass nicht direkt die Verteilung der UMC genutzt wird, sondern die Verteilung der Prognosefehler, welche die UMC beinhalten. Hierfür sind die folgenden Gleichungen des KF rekursiv für $t = 1, \dots, T$ zu durchlaufen (vgl. Abschnitt 2.3 und z. B. Harvey 1989, S. 105 ff.):

$$\mathbf{a}_{t|t-1} = \mathbf{a}_{t-1|t-1}, \quad \mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \mathbf{Q}_t, \quad (3.26)$$

$$\mathbf{e}_t = \mathbf{y}_t - \mathbf{H}_t \mathbf{a}_{t|t-1} - \mathbf{d}_t, \quad \boldsymbol{\Sigma}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \mathbf{R}_t, \quad (3.27)$$

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{H}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{e}_t, \quad \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{H}_t \mathbf{P}_{t|t-1}. \quad (3.28)$$

Die für den Start der Rekursion notwendigen Werte ($\mathbf{a}_{0|0} = [\bar{\boldsymbol{\alpha}}_0 \ \bar{\boldsymbol{\beta}}_0]'$) werden als konstante Parameter mitgeschätzt und $\mathbf{P}_{0|0}$ ist in diesem Fall eine $(K+1) \times (K+1)$ -Matrix mit Nullen (Harvey 1989, S. 137).¹⁴ Die restlichen Vektoren und Matrizen entsprechen den in Abschnitt 3.2.4 erläuterten Definitionen für das ZRM. Es sei in Erinnerung gerufen, dass \mathbf{y}_t den Durchschnittsnutzen $\boldsymbol{\delta}_t$ aus der vorherigen Marktanteilsinvertierung enthält.

¹⁴ Vgl. hierzu ebenfalls Abschnitt 2.3 im 2. Kapitel der vorliegenden Arbeit.

Mit dem *Prognosefehler* \mathbf{e}_t aus der KF-Rekursion sowie dessen Kovarianzmatrix $\mathbf{\Sigma}_t$ lässt sich die Likelihoodfunktion des Nachfragemodells aufstellen:¹⁵

$$\mathcal{L}^N = \ln L(\boldsymbol{\theta}^N | \boldsymbol{\delta}, \mathbf{p}, \mathbf{x}, \boldsymbol{\zeta}^*) = \ln \left(\prod_{t=1}^T f(\boldsymbol{\delta}_t, \mathbf{p}_t, \mathbf{x}_t, \boldsymbol{\zeta}_t^* | \boldsymbol{\theta}^N) \right) \quad (3.29)$$

$$= -0,5 \cdot \sum_{t=1}^T \left[\ln(2\pi^J |\mathbf{\Sigma}_t|) + \mathbf{e}_t' \mathbf{\Sigma}_t^{-1} \mathbf{e}_t \right]. \quad (3.30)$$

Der Vektor $\boldsymbol{\theta}^N = [\boldsymbol{\rho} \ \boldsymbol{\sigma}_\varepsilon \ \text{vec}(\mathbf{\Psi}) \ \mathbf{a}_{0|0} \ \text{vec}(\mathbf{Q})]'$ beinhaltet alle zu schätzenden konstanten Parameter des Nachfragemodells.

Berechnung der Jacobi-Determinante

Für die Aufstellung der Log-Likelihoodfunktion ist eine *nichtlineare Transformation* (Logitmodell) der Marktanteile notwendig. Aus dem *Gesetz zum Austausch von Variablen in Dichtefunktionen* (Judge *et al.* 1988, S. 30 ff.) folgt, dass daher die *Determinante der Jacobimatrix* der nichtlinearen Transformation $|\mathbf{J}_{\mathbf{e} \rightarrow \mathbf{s}}|$ zu berechnen ist (Jiang *et al.* 2009, S. 137 f.). $\mathbf{J}_{\mathbf{e} \rightarrow \mathbf{s}}$ bezeichnet die Matrix der ersten Ableitungen der Prognosefehler nach den Marktanteilen. Hierbei handelt es sich um eine blockdiagonale Matrix mit T Blöcken, die jeweils die Dimension $J \times J$ haben. Eine Vernachlässigung dieses Terms führt zu einer *Fehlspezifizierung der gesamten Log-Likelihoodfunktion* und damit zu Parameterverzerrungen. Wie Jiang *et al.* (2009) allerdings zeigen, hängt die Jacobimatrix jedoch nur von den *Heterogenitätsparametern* ab und somit hat ihre Vernachlässigung in homogenen Logitmodellen für die Maximierung der Log-Likelihoodfunktion keine Konsequenzen.¹⁶

Über die Beziehung $|\mathbf{J}_{\mathbf{e} \rightarrow \mathbf{s}}| = |\mathbf{J}_{\mathbf{s} \rightarrow \mathbf{e}}|^{-1}$ ist die Berechnung der Jacobi-Determinante leicht nachzuvollziehen. Es gilt $\forall t$ (Jiang *et al.* 2009, S. 138):

$$|\mathbf{J}_{\mathbf{s}_t \rightarrow \mathbf{e}_t}| = \left| \begin{bmatrix} \partial s_{1t} / \partial e_{1t} & \cdots & \partial s_{1t} / \partial e_{Jt} \\ \vdots & \ddots & \vdots \\ \partial s_{Jt} / \partial e_{1t} & \cdots & \partial s_{Jt} / \partial e_{Jt} \end{bmatrix} \right|. \quad (3.31)$$

Weil der Prognosefehler linear in $\boldsymbol{\delta}_t$ ist, gilt für die einzelnen Ableitungsterme:

¹⁵ Zwecks besserer Übersichtlichkeit in den Formeln unterbleibt die explizite Darstellung sämtlicher Abhängigkeiten. Diese sind aber aus Abbildung 3.1 ersichtlich.

¹⁶ Vgl. hierzu Abschnitt 2.5 der vorliegenden Arbeit.

$$\frac{\partial s_{jt}}{\partial e_{kt}} = \begin{cases} \int \text{pr}_{ijt}(1 - \text{pr}_{ijt})\phi(\mathbf{v}_i|\mathbf{0}, \mathbf{\Psi})d\mathbf{v}_i & j = k, \\ - \int \text{pr}_{ijt}\text{pr}_{ikt}\phi(\mathbf{v}_i|\mathbf{0}, \mathbf{\Psi})d\mathbf{v}_i & j \neq k. \end{cases} \quad (3.32)$$

Die Berechnung der Integrale in (3.32) erfordert ebenfalls Monte Carlo-Techniken, da keine geschlossenen Lösungen existieren. Praktischerweise liegen aber bereits simulierte Werte für die individuellen Wahlwahrscheinlichkeiten (pr_{ijt}) aus der Marktanteilsinvertierung vor, die hierfür eingesetzt werden können.

Die Likelihoodfunktion des gesamten Modells

Die zuvor beschriebenen vier Schritte liefern alle notwendigen Bestandteile, um die Likelihoodfunktion des gesamten Modells aufzustellen. Es ist das Produkt der Likelihoodfunktionen des Nachfrage- und Preismodells sowie der Jacobi-Determinante zu bilden. Für die Log-Likelihoodfunktion gilt dann entsprechend (Jiang *et al.* 2009, S. 145):

$$\mathcal{L}(\boldsymbol{\theta}^N, \boldsymbol{\theta}^P | \mathbf{s}, \mathbf{x}, \mathbf{p}, \mathbf{z}) = \mathcal{L}^N + \mathcal{L}^P - \ln(|\mathbf{J}_{\mathbf{s}_t \rightarrow \mathbf{e}_t}|). \quad (3.33)$$

Aus der Maximierung dieser Likelihoodfunktion folgt der MSL-Schätzer für alle konstanten (in $\boldsymbol{\theta}$ zusammengefassten) Modellparameter:

$$\boldsymbol{\theta}_{MSL} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \mathcal{L}(\boldsymbol{\theta}^N, \boldsymbol{\theta}^P | \mathbf{s}, \mathbf{x}, \mathbf{p}, \mathbf{z}). \quad (3.34)$$

Für (3.34) existiert keine geschlossene Lösung, daher kommt mit dem *Broyden-Fletcher-Goldfarb-Shanno*-Algorithmus ein gradientenbasiertes Verfahren für die numerische Maximierung der Log-Likelihoodfunktion zum Einsatz (Greene 2011, S. 1099). Pro Iteration sind jeweils alle zuvor beschriebenen und in Abbildung 3.1 skizzierten Schritte zu durchlaufen.

3.3.2 Zeitvariable Parameter

Zur Herleitung der Likelihoodfunktion des Nachfragemodells wurde die KF-Rekursion bereits vorgestellt. Dort war sie lediglich notwendig, um die Prognosefehler zu berechnen. Die Schätzung der zeitvariablen Parameter erfolgte quasi als Nebenprodukt. Hierbei wird zum Zeitpunkt t

aber immer nur die Information genutzt, welche *bis einschließlich Periode t* zur Verfügung steht. Dies ist für *prädiktive Anwendungen* von Vorteil: Nach der Schätzung der konstanten Parameter lässt sich der KF pro Periode anwenden, und sobald neue Daten vorliegen, findet ein *Update* der zeitvariablen Parameter statt. Der Modell-Anwender (z. B. ein Händler) bekommt stets datenbasiert und adaptiv die „aktuellen“ Nutzenparameter zur Nachfragemodellierung.

Soll aber die *gesamte* in den Daten *vorhandene Information* (also bis einschließlich T) in die Schätzung sämtlicher Parameterwerte für $t = 1, \dots, T$ einfließen, weil bspw. *retrospektiv* die Entwicklung der zeitvariablen Parameter untersucht wird, ist die Anwendung des KS notwendig. Hierbei handelt es sich ebenfalls um einen rekursiven Algorithmus, der für $t = T - 1, T - 2, \dots, 1$ die Resultate des KF verarbeitet (Petrís *et al.* 2009, S. 61):

$$\mathbf{a}_{t|T} = \mathbf{a}_{t|t} + \mathbf{G}_t(\mathbf{a}_{t+1|T} - \mathbf{a}_{t+1|t}) \text{ und} \quad (3.35)$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{G}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{G}_t'. \quad (3.36)$$

Die $(K + 1) \times (K + 1)$ -Matrix \mathbf{G}_t , mit $\mathbf{G}_t = \mathbf{P}_{t|t}\mathbf{P}_{t+1|t}^{-1}$, bestimmt, wie stark die Korrektur ausfällt. Ziel dieses Vorgehens ist es, mit der zusätzlichen Information eine *Glättung* der Parameterverläufe des KF vorzunehmen.

Nachdem also die Schätzung der konstanten Parameter abgeschlossen ist, werden die Ergebnisse des KF aus der letzten Iteration in den KS eingesetzt, um die geglätteten Schätzungen der zeitvariablen Parameter zu erhalten (Petrís *et al.* 2009, S. 144 ff.).

3.4 Simulationsexperiment

In diesem Abschnitt werden Aufbau und Ergebnisse einer *Monte Carlo-Simulation* präsentiert.¹⁷ Die Durchführung eines Simulationsexperiments im Rahmen des vorliegenden Beitrags lässt sich mehrfach begründen:

- *Ergründen statistischer Eigenschaften:* Im Simulationsexperiment sind das *wahre Modell*, die *wahren Fehlerterme* und vor allem die *wahren Parameter* bekannt. Folglich ist es möglich, zu evaluieren, wie gut

¹⁷ Einen umfassenden Überblick zu Simulationsexperimenten liefert z. B. Kiviet (2011).

unterschiedliche Modellspezifikationen in der Lage sind, die wahren Parameter zu ermitteln, bzw. wie groß etwaige Verzerrungen ausfallen (Shugan 2002). Zudem wird ersichtlich, wie präzise die Schätzergebnisse ausfallen.

- *Verbessern des Ergebnisverständnisses:* Monte Carlo-Simulationen helfen zudem, empirische Ergebnisse besser zu *verstehen* und zu *interpretieren* (Gelman und Hill 2007, S. 157 f.). Das gilt vor allem für Ergebnisse, die keinen „Lehrbuchcharakter“ besitzen.
- *Testen der Software:* Es ist *keine Standardsoftware* für die hier untersuchten Modelle erhältlich, weshalb entsprechende Programme zu erstellen sind.¹⁸ Aus diesem Grund ist die Simulation ein unverzichtbares Werkzeug, um die Richtigkeit der Programme zu überprüfen und zu bestätigen (Cook *et al.* 2006).

Beginnend mit einem *datengenerierenden Prozess* (DGP), der auf dem präsentierten Logitmodell mit *Heterogenität*, *Endogenität* und *Dynamik* fußt, werden Datensätze simuliert und anschließend jeweils mit *drei verschiedenen Modellspezifikationen* geschätzt. Beim ersten Modell (TVP-MXL (CF))¹⁹ handelt es sich um das Modell, welches zum DGP passt und den CF-Ansatz zur Endogenitätskorrektur einsetzt. Modell 2 (TVP-MXL (IV)) berücksichtigt zwar *Heterogenität*, *Endogenität* und *Dynamik*, *ad hoc* werden hier aber fälschlicherweise endogenitätsbereinigte Preise anstatt der CF eingesetzt (siehe Gleichungen 3.19 und 3.20).²⁰ Bei Modell 3 (TVP-MXL) wird gänzlich auf eine Berücksichtigung der Endogenität verzichtet. Die Modelle 2 und 3 sind demnach *fehlspezifiziert* und die Auswahl dieser Modellspezifikationen dient dazu, *drei Fragen* zu beantworten, welche Tabelle 3.1 zu entnehmen sind.

3.4.1 Datengenerierender Prozess

Der DGP für die Simulation fußt auf dem diskreten Wahlmodell für aggregierte Daten aus Abschnitt 3.2. Inhaltlich sollen aus Händlersicht sowohl Nachfrage als auch Preise für Marken einer typischen Kategorie schnell-drehender Konsumgüter nachempfunden werden. Der Aufbau orientiert

¹⁸ Im vorliegenden Beitrag wird die freie Software R (R Core Team 2013) verwendet.

¹⁹ „TVP“ steht für *time varying parameters*.

²⁰ Die Modellbezeichnung TVP-MXL (IV) soll andeuten, dass, wie bei typischen Instrumentenvariablen-Schätzungen, eine Korrektur der endogenen Variable stattfindet.

Tabelle 3.1: Modellübersicht und Forschungsfragen

Modell	Endogenität	Forschungsfrage
TVP-MXL (CF)	✓	Sind mit dem richtigen Modell die wahren Parameter zu identifizieren?
TVP-MXL (IV)	⚡	Welche Konsequenzen resultieren, wenn Endogenität falsch berücksichtigt wird?
TVP-MXL	—	Welche Konsequenzen resultieren, wenn Endogenität vernachlässigt wird?

sich stark an Jiang *et al.* (2009, S. 140 ff.) und der empirischen Studie in Abschnitt 3.5.

Es gibt drei Marken ($J = 3$) sowie eine Nichtwahloption und die Nachfrage wird nur über *ein* (endogenes) Marketinginstrument beeinflusst, den *logarithmierten Preis*. Als (exogenes) Instrument steht der *logarithmierte Großhandelspreis* zur Verfügung, welcher aus Sicht des Händlers vereinfachend als variable Kosten interpretiert werden kann. Diese logarithmierte Variable wird als gleichverteilte Zufallsgröße simuliert, mit Werten im Intervall $[-1,000, -0,200]$. Dies impliziert Großhandelspreise im Intervall $[0,368, 0,819]$ mit einem Mittelwert von 0,56 Geldeinheiten (z. B. €) mit einer Standardabweichung von 0,13. Zur Begünstigung der Identifizierung zeitvariabler Parameter gibt es mit $T = 300$ (Wochen) einen langen Zeithorizont (Sriram und Kalwani 2007).

Folgende *Parametrisierung* liegt der Simulationsstudie zugrunde. Im Preismodell wird für die drei Markenkonstanten und den über alle Marken einheitlichen Kostenparameter $[\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_{\text{Kosten}}]' = [0,2 \ 0,1 \ 0,1 \ 0,6]'$ angenommen. Der Wert des Kostenparameters von 0,6 ist wegen der logarithmierten Preise und Kosten als *Pass-Through-Elastizität*²¹ zu interpretieren und besitzt einen realistischen Wert (vgl. z. B. Nijs *et al.* 2010).

Für die einzelnen Elemente der *Kovarianzmatrix* Ω gilt $\{\sigma_{\varepsilon j}\}_{j=1}^J = 0,5$, $\{\sigma_{\zeta j}\}_{j=1}^J = 0,2$ und $\{\rho_j\}_{j=1}^J = 0,4$. Für das Preismodell implizieren die $\sigma_{\zeta j}$ -Werte im Zusammenhang mit der Variation der Kostenvariable, dass ca. 60 % der Preisvariation ($R^2 = 0,6$) erklärt werden kann und somit kein Problem schwacher Instrumente vorliegt. Der Korrelationswert von 0,4 entspricht zudem in etwa der Annahme von Jiang *et al.* (2009) und

²¹ *Pass-Through* ist definiert als der Anteil einer Änderung im Großhandelspreis, der vom Händler an den Endverbraucher weitergegeben wird.

wird von den Autoren als *moderates Maß an Endogenität* interpretiert. Wegen des positiven Vorzeichens gehen positive Preisschocks mit positiven Nutzen- bzw. Nachfrageschocks einher. Ohne eine Endogenitätskorrektur fallen die Preiseffekte hierdurch absolut gesehen niedriger aus.²²

$[\bar{\alpha}_{10} \ \bar{\alpha}_{20} \ \bar{\alpha}_{30} \ \bar{\beta}_0]' = [-5 \ -6 \ -6 \ -5]'$ gilt für die *Startwerte* der *Parameter*. Für die Interpretation der Werte der $\bar{\alpha}_{j0}$ -Parameter sind nur deren Differenzen zueinander relevant und folglich besitzt hier die Marke $j = 1$ den größten Nutzen. Da die Nichtwahloption einen Nutzen von Null hat, spiegeln die negativen Werte lediglich wider, dass die Nichtwahloption einen recht hohen Marktanteil ($> 90\%$) besitzt. Die zeitliche Variation der Parameter basiert auf den Standardabweichungen $\{\sigma_{\eta_{\alpha j}}\}_{j=1}^J = 0,05$ sowie $\sigma_{\eta_{\beta}} = 0,1$. Auf diese Weise nehmen die Parameter und damit die Nutzen für alle Perioden plausible Werte an.²³ Der Einfachheit halber wird *Heterogenität* nur im Preiseffekt unterstellt mit $\sigma_{v_{\beta}}^2 = 2$.

Die Simulation und Schätzung der Preisparameterheterogenität erfolgt über $S = 250$ Zufallsziehungen aus $N(0,1)$ basierend auf Haltonsequenzen. Es werden $\mathcal{R} = 100$ verschiedene *Datensätze (Replikationen)* gemäß des beschriebenen DGPs erzeugt, indem aus den spezifizierten Verteilungen für das Nachfrage- und Preismodell sowie der zeitvariablen Parameter entsprechende Zufallszahlen gezogen werden. Jede Replikation basiert demnach nicht nur auf unterschiedlichen Realisationen der Störgrößen, sondern ebenfalls auf verschiedenen Verläufen für die zeitvariablen Parameter (Kim und Kim 2011, S. 493). Die exogene Instrumentalvariable wird nicht neu gezogen. Für jede Replikation wird die zuvor vorgestellte MSL-Schätzung durchgeführt und die zeitvariablen Parameter werden mit dem KS bestimmt.

3.4.2 Ergebnisse der Simulation

Konstante Parameter

Zunächst werden die Ergebnisse der *konstanten* Parameter diskutiert. In Tabelle 3.2 sind für die jeweiligen Parameter der drei Modelle die *Mit-*

²² Eine Diskussion bzgl. der Richtung der Endogenitätsverzerrung ist in Train (2009, S. 316 f.) sowie Park und Gupta (2009, S. 540) zu finden. Alle Simulationen wurden ebenfalls mit negativer Korrelation ($\rho = -0,4$) durchgeführt, wobei dies keinen substantziellen Einfluss auf die Ergebnisse hatte.

²³ Insbesondere sollte der Preiseffekt stets negativ bleiben. Garantiert werden kann das bei Random Walk-Prozessen zwar nicht, die gewählte Parametrisierung begünstigt dies aber.

telwerte und die Standardabweichungen über alle Replikationen abgetragen. Zudem sind der Tabelle die wahren Parameterwerte zu entnehmen sowie die Ergebnisse für t -Tests der Simulationsergebnisse gegen die wahren Parameterwerte.²⁴

Tabelle 3.2: Resultate des Monte Carlo-Experiments (konst. Parameter)

Parameter	wahrer Wert	TVP-MXL (CF)		TVP-MXL (IV)		TVP-MXL	
		Mittelwert	Std-Abw.	Mittelwert	Std-Abw.	Mittelwert	Std-Abw.
γ_1	0,200	0,201	0,024	0,201	0,023	0,201	0,024
γ_2	0,100	0,101	0,021	0,102	0,021	0,101	0,021
γ_3	0,100	0,100	0,022	0,101	0,021	0,100	0,022
γ_{Kosten}	0,600	0,604	0,032	0,604	0,032	0,602	0,032
σ_ζ	0,200	0,199	0,005	0,199	0,005	0,199	0,005
$\tilde{\alpha}_{10}$	-5,000	-5,039*	0,164	-5,032	0,245	-4,922*	0,164
$\tilde{\alpha}_{20}$	-6,000	-6,018	0,145	-5,995	0,257	-5,855*	0,147
$\tilde{\alpha}_{30}$	-6,000	-5,988	0,169	-5,960	0,272	-5,820*	0,170
$\tilde{\beta}_0$	-5,000	-5,031	0,357	-5,084	0,694	-4,360*	0,355
σ_ε	0,500	0,502	0,063	0,981*	0,166	0,475*	0,012
ρ	0,400	0,396	0,056				
σ_{η_α}	0,050	0,047*	0,008	0,043*	0,016	0,047*	0,008
σ_{η_β}	0,100	0,097	0,021	0,073*	0,061	0,097	0,022
σ_{ν_β}	1,414	1,388	0,173	0,890*	0,761	1,436	0,179

Anmerkung: Die Ergebnisse basieren auf $\mathcal{R} = 100$ Replikationen und θ_r fasst die Ergebnisse für die r -te Replikation aller zeitlich konstanten Parameter eines Modells zusammen. Ferner bezeichnet θ^w den Vektor der wahren Parameterwerte. Es gilt: Mittelwert: $\bar{\theta} = (1/\mathcal{R}) \cdot \sum_{r=1}^{\mathcal{R}} \theta_r$ und Std-Abw.: $\sigma_\theta = ((1/(\mathcal{R}-1)) \cdot \sum_{r=1}^{\mathcal{R}} (\theta_r - \bar{\theta})^2)^{1/2}$. „*“ bedeutet, dass der t -Wert den kritischen Wert auf dem 5 % Niveau ($|t| > 1,984$) überschreitet, mit $t = \mathcal{R}^{1/2} \cdot (\bar{\theta} - \theta^w) \cdot \sigma_\theta^{-1}$.

- *Preismodell*: Die Ergebnisse für die Parameter des Preismodells fallen bei allen Modellen durchgängig sehr gut aus. Für das TVP-MXL ist das Preismodell zwar irrelevant, es wurde hier aber trotzdem mitgeschätzt, um zu ermitteln, welche Ergebnisse bei einer isolierten Untersuchung des Preismodells resultieren. Hierdurch wird ersichtlich, dass die Wahl des Nachfragemodells keinen Einfluss (bis auf Rundungsfehler) auf die Ergebnisse des Preismodells hat, was nicht weiter überraschend ist, da es sich um ein deskriptives Preismodell handelt. Dieser Schritt dient folglich der Absicherung, dass die Simulation

²⁴ Vgl. für ein ähnliches Vorgehen zur Auswertung von Simulationsexperimenten z. B. Fiebig *et al.* (2010).

fehlerfrei arbeitet.

- *TVP-MXL (CF)-Modell:* Bei Verwendung des richtigen Modells sind die Schätzergebnisse sehr zufriedenstellend. Die Mittelwerte liegen nah an den wahren Parameterwerten und die empirischen Standardabweichungen sind klein. Lediglich bei α_{10} und $\sigma_{\eta\alpha}$ ist eine signifikante Verzerrung vorhanden. Für $\sigma_{\eta\alpha}$ ist als Grund anzuführen, dass es u. U. (empirisch) herausfordernd ist, zeitvariable Grundnutzen zu identifizieren und deren Effekt von dem des „normalen Störterms“ zu trennen. Die Verzerrungen liegen jedoch nur im niedrigen einstelligen Prozentbereich und sind demnach nicht weiter besorgniserregend. Von besonderem Interesse ist $\sigma_{v\beta}$. Die Simulation zeigt, dass es gut möglich ist, Heterogenität aus aggregierten Daten zu schätzen. Dabei ist die Präzision der Schätzung durchaus vergleichbar mit dem likelihoodbasierten Schätzer von Park und Gupta (2009).
- *TVP-MXL (IV)-Modell:* Im dynamischen Modell mit falscher Endogenitätsberücksichtigung ist festzustellen, dass sämtliche Verteilungsparameter Verzerrungen aufweisen. Auffällig ist zudem, dass die Schätzungen im Nachfragemodell deutlich unpräziser sind als im TVP-MXL (CF)-Modell; teilweise sind die empirischen Standardabweichungen bis zu fünfmal größer (z. B. für $\sigma_{v\beta}$). σ_ε wird extrem überschätzt. Dies ist einleuchtend, weil der mit den UMC korrelierte Preisbestandteil vom Störterm aufgenommen wird, weshalb σ_ε zu groß ausfällt. Die genauen Konsequenzen hiervon auf die Ergebnisse der zeitvariablen Parameter werden nachfolgend ausführlich diskutiert. Die Preisparameterheterogenität wird im Durchschnitt unterschätzt. Jedoch belegt die sehr große Standardabweichung, dass die Schätzungen sehr unpräzise sind und z. T. auch deutlich über dem wahren Wert liegen.
- *TVP-MXL-Modell:* Ohne Endogenitätsberücksichtigung sind die Ergebnisse bzgl. Heterogenität und Dynamik ($\sigma_{\eta\alpha}$, $\sigma_{\eta\beta}$ und $\sigma_{v\beta}$) vergleichbar mit dem TVP-MXL (CF)-Modell. Die Startwerte der Nutzenparameter ($\{\alpha_{j0}\}_{j=1}^J$ und β_0) sind dagegen verzerrt. β_0 wird z. B. (absolut betrachtet) unterschätzt, was auf das positive Vorzeichen von ρ zurückzuführen ist. Interessanterweise sind die Schätzungen alles in allem ähnlich effizient, wie beim TVP-MXL (CF)-Modell. Es bleibt damit festzuhalten, dass das Ignorieren von Preisendogenität zu typischen Verzerrungen im Nachfragemodell führt. Negative Konsequenzen auf die Schätzung von Heterogenität und Dynamik

scheint es jedoch nicht zu geben.

Zeitvariable Parameter

Nach der Untersuchung der Ergebnisse für die konstanten Parameter liegt der Fokus nun auf den Ergebnissen bzgl. der *zeitvariablen* Parameter, welche in Tabelle 3.3 dargestellt sind. Um die Qualität der Schätzung der zeitvariablen Parameter aussagekräftig beurteilen zu können, wird für jede Replikation berechnet, wie hoch der *mittlere Fehler* (me), die *Wurzel des mittleren quadratischen Fehlers* (rmse), die *Korrelation* (cor) sowie die *maximale absolute Abweichung* (mxad) pro Parameter über die Zeit ausfällt und anschließend über alle Replikationen gemittelt.

- *TVP-MXL (CF)-Modell*: Das richtige Modell ist in der Lage, die zeitvariablen Parameter präzise zu schätzen. Die wahren und die geschätzten Parameterverläufe weichen nicht nennenswert voneinander ab, die Korrelationen fallen sehr hoch aus und die maximalen absoluten Abweichungen besitzen mit 0,714 ein vertretbares Maß.
- *TVP-MXL (IV)-Modell*: Im Falle einer falschen Berücksichtigung der Endogenität sind die geschätzten Parameterverläufe im Durchschnitt leicht verzerrt, was aus den Werten für den mittlere Fehler ersichtlich ist. Die Markenkonzanten (α_{jt}) werden im Mittel überschätzt, der zeitvariable Preisparameter (β_t) fällt dagegen (absolut gesehen) zu hoch aus. Auffällig ist, dass die geschätzten Verläufe deutlich weniger mit den wahren Parameterverläufen korrelieren als das TVP-MXL (CF)-Modell. Zudem fallen auch die maximalen absoluten Abweichungen größer aus. Auch wenn die durchschnittlichen Verzerrungen nicht besonders hoch ausfallen, lassen die Ergebnisse den Schluss zu, dass die Schätzungen der Parameterverläufe im Gegensatz zum TVP-MXL (CF)-Modell schlechter ausfallen.
- *TVP-MXL-Modell*: Das TVP-MXL-Modell zeigt ein umgekehrtes Bild. Hier sind klare durchschnittliche Verzerrungen feststellbar, die ebenfalls mit den Abweichungen der Startwerte in Tabelle 3.2 korrespondieren. Vor allem der Preisparameter wird unterschätzt und die Endogenitätsproblematik ist evident. Die Korrelationen liegen dagegen auf einem ähnlichen Niveau wie beim TVP-MXL (CF)-Modell. Im Vergleich zum TVP-MXL (IV)-Modell sind die maximalen absoluten Abweichungen trotz Endogenitätsverzerrungen sogar etwas niedriger.

Tabelle 3.3: Resultate des Monte Carlo-Experiments (zeitvar. Parameter)

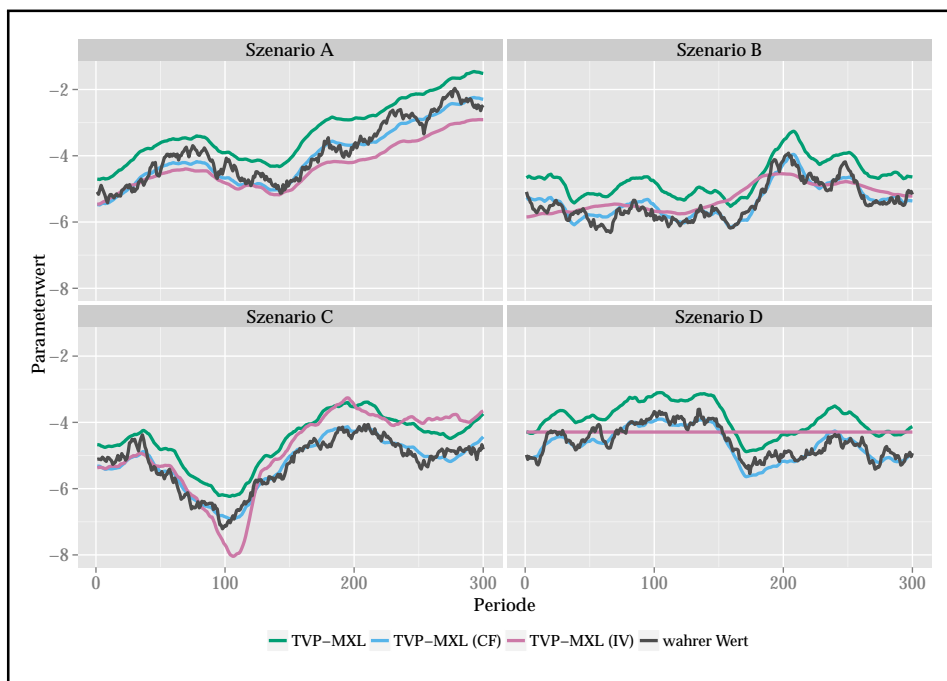
Para- meter	TVP-MXL (CF)				TVP-MXL (IV)				TVP-MXL			
	me	rmse	cor	mxad	me	rmse	cor	mxad	me	rmse	cor	mxad
$\tilde{\alpha}_{1t}$	-0,004	0,109	0,910	0,322	0,026	0,165	0,807	0,451	0,106	0,110	0,907	0,386
$\tilde{\alpha}_{2t}$	0,002	0,111	0,899	0,333	0,038	0,177	0,751	0,494	0,168	0,113	0,897	0,456
$\tilde{\alpha}_{3t}$	0,004	0,115	0,920	0,335	0,038	0,181	0,814	0,492	0,178	0,116	0,917	0,475
$\tilde{\beta}_t$	0,000	0,222	0,921	0,714	-0,039	0,453	0,697	1,291	0,687	0,255	0,920	1,259

Anmerkung: Die Ergebnisse basieren auf $\mathcal{R} = 100$ Replikationen mit jeweils $T = 300$ Perioden und $\boldsymbol{\theta}_{t,r}$ fasst die Ergebnisse für die r -te Replikation und t -te Periode aller zeitvariablen Parameter eines Modells zusammen. Ferner bezeichnet $\boldsymbol{\theta}_{t,r}^w$ den Vektor der wahren Parameterwerte. Alle Gütenmaße werden pro Replikation in der Zeitdimension berechnet und dann über alle Replikationen gemittelt. Es gilt: Mittlerer Fehler (me) = $(1/\mathcal{R}) \cdot \sum_{r=1}^{\mathcal{R}} (1/T) \cdot \sum_{t=1}^T (\boldsymbol{\theta}_{t,r} - \boldsymbol{\theta}_{t,r}^w)$; Wurzel des mittleren quadrierten Fehlers ($rmse$) = $(1/\mathcal{R}) \cdot \sum_{r=1}^{\mathcal{R}} ((1/T) \cdot \sum_{t=1}^T (\boldsymbol{\theta}_{t,r} - \boldsymbol{\theta}_{t,r}^w)^2)^{1/2}$; Korrelation (cor) = $(1/\mathcal{R}) \cdot \sum_{r=1}^{\mathcal{R}} ((1/T) \cdot \sum_{t=1}^T (\boldsymbol{\theta}_{t,r} - \boldsymbol{\theta}_{t,r}^w) \cdot (\boldsymbol{\theta}_{t,r}^w - \boldsymbol{\theta}_{t,r}^w)) / ((\sum_{t=1}^T (\boldsymbol{\theta}_{t,r} - \boldsymbol{\theta}_{t,r}^w)^2)^{1/2} \cdot (\sum_{t=1}^T (\boldsymbol{\theta}_{t,r}^w - \boldsymbol{\theta}_{t,r}^w)^2)^{1/2})$; Maximale absolute Abweichung ($mxad$) = $(1/\mathcal{R}) \cdot \sum_{r=1}^{\mathcal{R}} \max_t (\text{abs}(\boldsymbol{\theta}_{t,r} - \boldsymbol{\theta}_{t,r}^w))$.

Zusammenfassend fällt auf, dass die Verwendung des richtigen Modells eine sehr gute Schätzung der zeitvariablen Parameter ermöglicht. Eine falsche oder fehlende Behandlung der Preisendogenität führt dagegen zu einer Verzerrung der Parameter (insb. des Preisparameters). Die Verzerrungen sind allerdings unterschiedlicher Natur. Die Ergebnisse lassen vermuten, dass das Ignorieren der Endogenität „lediglich“ zu einer *Parallelverschiebung* der Parameter führt, die Gestalt der Verläufe wird jedoch gut abgebildet. Beim TVP-MXL (IV)-Modell dagegen kommt es ebenfalls zu *Verzerrungen in den Verläufen*, im Durchschnitt werden die wahren Parameter-niveaus aber hinreichend gut getroffen.

Um diese Erkenntnisse näher zu untersuchen, erfolgt eine *visuelle Inspektion* der Preisparameterverläufe *aller Replikationen*. Es gibt eine Reihe ähnlicher Resultate, die sich grob in vier *Szenarien* (A bis D) einteilen lassen. Abbildung 3.2 zeigt vier Replikationen, die jeweils stellvertretend für ein Szenario sind.

Abbildung 3.2: Grafische Darstellung der zeitlichen Preisparameterverläufe ausgewählter Replikationen



In allen Szenarien sind die Erkenntnisse, die auf Basis der Gütemaße in Tabelle 3.3 gewonnen wurden, direkt zu erkennen: 1.) Mit dem TVP-MXL (CF)-Modell lassen sich die wahren Preisparameterverläufe treffge-

nau schätzen und 2.) gibt es beim TVP-MXL-Modell Verzerrungen in Form einer Parallelverschiebung. Für das TVP-MXL (IV)-Modell fallen die Resultate für die vier Szenarien differenzierter aus und hängen vom jeweiligen Datensatz ab. Im besten Fall (*Szenario A*) sind die Unterschiede in den Ergebnissen zwischen dem TVP-MXL (IV)- und dem TVP-MXL (CF)-Modell gering. Die Überlegenheit des TVP-MXL (CF)-Modells ist trotzdem offensichtlich. In *Szenario B* fällt die zeitliche Variation in den geschätzten Preisparametern des TVP-MXL (IV)-Modells zu gering aus, eine gewisse Ähnlichkeit zum wahren Parameterverlauf ist jedoch vorhanden. In diesem Szenario ist zu vermuten, dass die Verzerrungen des TVP-MXL (IV)-Modells weniger problematisch sind als im TVP-MXL-Modell. Es kann jedoch auch sein (*Szenario C*), dass es eine sehr hohe zeitliche Variation in den geschätzten Preisparametern des TVP-MXL (IV)-Modells gibt. Hier besteht die Möglichkeit, dass das TVP-MXL-Modell weniger Verzerrungen aufweist als das TVP-MXL (IV)-Modell. In *Szenario D* wird vom TVP-MXL (IV)-Modell keine Dynamik im Preisparameter identifiziert, obwohl sie vorhanden ist und dies gilt für 30 der 100 Replikationen. Hier verfehlt das TVP-MXL (IV)-Modell folglich seinen Zweck und ist wertlos, wenn die Schätzung zeitvariabler Parameter im Vordergrund steht.

Als *Fazit* der Monte Carlo-Simulation lässt sich festhalten, dass für eine verlässliche Endogenitätsbehandlung im Falle zeitvariabler Parameter der vorgestellte CF-Ansatz erforderlich ist (TVP-MXL (CF)-Modell). Andernfalls können die Ergebnisse bei einer vermeintlichen Endogenitätskorrektur (TVP-MXL (IV)-Modell) schlechter ausfallen als bei einer Nichtbeachtung der Endogenität (TVP-MXL-Modell). Ungeachtet dessen geht die Verwendung des TVP-MXL (IV)-Modells mit einem Verlust an (statistischer) Effizienz einher. Vor dem Hintergrund der guten Anwendbarkeit des TVP-MXL (CF)-Modells und dessen durchweg überzeugender Ergebnisse, kann eindeutig empfohlen werden, dieses Modell in entsprechenden Situationen einzusetzen.

3.5 Empirische Studie

Dieser Abschnitt verdeutlicht die Modellanwendung anhand einer *empirischen Studie*. Nach einer Beschreibung der Daten werden die Schätzergebnisse präsentiert und diskutiert. Abschließend illustriert die Betrachtung deckungsbeitragsmaximaler Preise aus Händlersicht die ökonomischen Implikationen der Endogenitätsberücksichtigung.

3.5.1 Daten

Bei den Daten handelt es sich um *Handelspanel*daten aus der *Dominick's Finer Foods (DFF)*-Datenbank der Universität Chicago (Booth School of Business).²⁵ DFF ist mit mehr als 80 Geschäften einer der marktführenden Händler im Großraum Chicago. Verwendet werden wöchentliche Werte für die Variablen *Absatz*, *Preis*, *Promotion* und *Kosten* (Großhandelspreise) der Produktkategorie *Dosenthunfisch* im Zeitraum von September 1989 bis Mai 1997 (knapp 400 Wochen). Aufgrund *fehlender Werte* stehen allerdings nur 337 Wochen für die Auswertung zur Verfügung. Die Daten sind auf Markenebene über die Geschäfte aggregiert und es werden die vier umsatzstärksten Marken *Bumble Bee* (BB), *Chicken of the Sea* (CS), die Handelsmarke *Heritage House* (HH) und *Star Kist* (SK) in der Untersuchung betrachtet, welche für den Großteil des Kategorieabsatzes verantwortlich sind. Absätze (und Preise) beziehen sich auf Thunfisch mit einer einfachen Qualität („chunky light“, vgl. Meza und Sudhir 2006) in Dosen der Größe 6 oz (≈ 170 Gramm).

Die Daten weisen aus, dass pro Woche fast 2 Mio. Kunden bei DFF einkaufen (Mittelwert = 1.934.321, Standardabweichung = 178.550). Diese wöchentliche Besucherzahl wird zur Definition des Marktvolumens (M_t) herangezogen, um damit die Marktanteile der vier Marken und der Nichtwahloption berechnen zu können. Diese Berechnung ist unter der Annahme möglich, dass sich die Wahlentscheidungen auf einzelne Dosen beziehen (Jiang *et al.* 2009, S. 142).

Tabelle 3.4 zeigt deskriptive Statistiken zu den DFF-Daten. SK besitzt den höchsten *Absatz* und damit auch den höchsten *Marktanteil*, dagegen hat die Handelsmarke von DFF (HH) den niedrigsten Absatz und den kleinsten Marktanteil. BB und CS liegen ungefähr auf gleichem Niveau und sind zwischen HH und SK einzuordnen. Die *Preise* betragen für CS und SK ca. 0,8 \$ pro Dose, BB und HH sind im Durchschnitt 0,02 \$ bzw. 0,11 \$ billiger. Ein ähnliches Bild ergibt sich für die *Kosten*, weshalb für DFF in dieser Produktkategorie eine *Deckungsbeitragsrate* im Durchschnitt von knapp über 29 % resultiert. Die Marken unterscheiden sich leicht in der Intensität der *Promotionaktivität*, wobei hier CS den höchsten Wert aufweist, gefolgt von SK sowie BB und HH.

Alle Variablen weisen eine *deutliche Variation über die Zeit* auf, was durch-

²⁵ <http://research.chicagobooth.edu/marketing/databases/dominicks/index.aspx>.

Tabelle 3.4: Deskriptive Statistiken zu den Thunfischdaten

Variable	Bumble Bee	Chicken of the Sea	Heritage House	Star Kist
Absatz	1,470 (3,323)*	1,443 (3,905)	0,924 (1,328)	2,085 (3,759)
Marktanteil	0,223 (0,186)	0,229 (0,186)	0,196 (0,137)	0,352 (0,205)
Preis	0,779 (0,112)	0,801 (0,085)	0,691 (0,068)	0,804 (0,102)
Kosten	0,544 (0,056)	0,573 (0,045)	0,497 (0,046)	0,561 (0,051)
Promotion	0,234 (0,348)	0,344 (0,405)	0,244 (0,364)	0,308 (0,389)

Anmerkung: (*) Mittelwert und Standardabweichung in Klammern. Die Variablen sind wie folgt skaliert: Absatz in 10^4 6 oz-Dosen, Marktanteile in 100 % sowie Preise und Kosten in US\$ pro 6 oz-Dose. Die Promotion-Variable bezieht sich auf den Prozentsatz des Absatzes, der unter Promotionbedingungen verkauft wurde. Bei den Marktanteilen handelt es sich um „Inside-Good Marktanteile“, sie sind demnach auf die Wahl einer Thunfischmarke bezogen. Das „Outside-Good“ selbst besitzt einen Marktanteil von ca. 97 %.

aus typisch für Daten auf Händlerebene ist und natürlich die Identifizierung von Marketing-Mix-Effekten begünstigt. Der DFF-Datensatz in der Produktkategorie Dosenthunfisch bietet vor dem Hintergrund des Untersuchungsgegenstands des vorliegenden Beitrags aber noch mehr Vorteile:

1. Die Daten erstrecken sich über einen *langen Zeitraum*, was die Identifizierung zeitvariabler Parameter erleichtert (Sriram und Kalwani 2007, S. 52 f.).
2. *Heterogenität* (Jiang *et al.* 2009), *Endogenität* (Meza und Sudhir 2006) und *Dynamik* in Form saisonaler Effekte (Chevalier *et al.* 2003) wurden bereits anhand dieses Datensatzes angesprochen. Eine gemeinsame Betrachtung der drei Aspekte mit einer flexiblen Modellierung zeitvariabler Parameter erscheint als logische Erweiterung der bestehenden Forschungsarbeiten.
3. Der Dosenthunfischmarkt lässt sich mit einer *geringen Anzahl Marken* adäquat abbilden, was den Analyseaufwand reduziert. BB, CS, HH und SK sind bereits für ca. 80 % des Marktes verantwortlich.
4. Dosenthunfisch wird *nicht stark beworben*, und daher ist es unproblematisch, dass hierzu keine Daten vorliegen. Bei anderen Produktkategorien, z. B. Bier oder Waschmitteln, ist dies nicht der Fall.

3.5.2 Instrumentalvariablen

Zur Verwendung des vorgestellten CF-Ansatzes sind brauchbare *Instrumentalvariablen* (z_{jt}) notwendig. Chintagunta *et al.* (2002), Meza und Sudhir (2006) sowie Sriram *et al.* (2007) folgend, wird der *Großhandelspreis* als Instrument für den Preis des Händlers herangezogen. Die Begründung hierfür lautet wie folgt (vgl. hierzu detailliert Chintagunta *et al.* 2002, S. 147 f.): Großhandelspreise sind i. d. R. hoch korreliert mit den Preisen des Händlers, weil für diesen die Großhandelspreise Kosten darstellen. Bei Kostenschwankungen passt der Händler seine Preise ggf. an.²⁶ Die UMC sind aber höchstwahrscheinlich nicht korreliert mit Kosten, da der Händler z. B. die Regalplatzierung nicht immer anpasst, wenn sich der Großhandelspreis ändert.

Neben diesem inhaltlichen Argument für Großhandelspreise als valides Instrument für endogene Händlerpreise existieren auch empirische Indizien. Nijs *et al.* (2007) untersuchen Determinanten für Händlerpreise und -gewinne und kommen zur Einsicht, dass Großhandelspreise hier eine maßgebliche Rolle spielen. Das gilt vor allem für die ebenfalls untersuchten DFF-Daten. Ferner führen Fong *et al.* (2011) in Kooperation mit einem Händler ein Feldexperiment durch, in welchem sie Preiselastizitäten ermitteln. Sie vergleichen ihre Ergebnisse mit Resultaten, die auf nicht-experimentellen Daten desselben Händlers fußen. Wie zu vermuten sind die geschätzten Elastizitäten aufgrund von Endogenität verzerrt. Der Einsatz von Großhandelspreisen als Instrument verhindert diese Verzerrung jedoch und die Autoren schlussfolgern: „*The wholesale price instrument eliminates far more bias than it introduces, and appears to be an effective way to estimate price sensitivity*“ (Fong *et al.* 2011, S. 17).

Auch die *Promotionvariable* wird als exogen angenommen und lässt sich folglich ebenfalls als Instrument verwenden. Dieses Vorgehen ist nicht unüblich im Marketing (vgl. z. B. Chintagunta 2001 und Sriram *et al.* 2007). Die Begründung hierfür lautet, dass Promotionaktivitäten üblicherweise mit mehreren Wochen Vorlauf geplant werden. Daher erscheint es unwahrscheinlich, dass Korrelationen zwischen der Promotionvariablen und dem Störterm auf Nachfrageebene vorliegen (Chintagunta *et al.* 2002). Promotionaktivitäten und Preise sind aber i. d. R. hoch korreliert, da Händler ihre Preise auf ihre geplanten Promotionaktivitäten abstimmen.

²⁶ Vgl. hierzu auch die Diskussion in Train (2009, S. 311). Mehr als 80 % der befragten Manager einer empirischen Studie geben an, dass sie einen festen Aufschlagsatz auf die Kosten zur Preisbildung („Kosten-Plus-Preisbildung“) nutzen.

Des Weiteren wird ein *Interaktionsterm* aus der *Kosten-* und *Promotionvariable* gebildet, um Pass-Through-Unterschiede in Promotionwochen zu erklären, und zur Abbildung der verschiedenen Preisniveaus der Marken (vgl. Tabelle 3.4) werden markenspezifische Konstanten spezifiziert. Da der Preis im Nachfragemodell logarithmiert wird, erfolgt ebenfalls eine Logarithmierung der Kosten. Zum einen führt das zu einem besseren Datenfit und zum anderen lassen sich die Resultate leichter auf Plausibilität prüfen, da in diesem Fall direkt die *Pass-Through-Elastizität* geschätzt wird.

Um einen Eindruck von der *Qualität der Instrumente* zu bekommen und deren Plausibilität zu überprüfen, wird zunächst eine gepoolte Regression der Preise auf die genannten Instrumentalvariablen durchgeführt. Gleichung (3.37) fasst die Ergebnisse zusammen (mit Standardfehlern in Klammern):

$$\begin{aligned} \ln(\text{Preis}_{jt}) = & \underset{(0,021)}{0,027} \cdot \mathbb{I}(j = \text{BB}) + \underset{(0,020)}{0,055} \cdot \mathbb{I}(j = \text{CS}) - \underset{(0,024)}{0,048} \cdot \mathbb{I}(j = \text{HH}) + \\ & \underset{(0,021)}{0,061} \cdot \mathbb{I}(j = \text{SK}) + \underset{(0,034)}{0,397} \cdot \ln(\text{Kosten}_{jt}) - \underset{(0,038)}{0,117} \cdot \text{Promo}_{jt} + \\ & \underset{(0,060)}{0,111} \cdot \ln(\text{Kosten}_{jt}) \cdot \text{Promo}_{jt}, \end{aligned} \quad (3.37)$$

$$R^2 = 0,548, N = 1348.$$

Mit einem R^2 -Wert von 0,548 besitzen die Instrumente eine hinreichende Erklärungskraft. Die Kosten haben einen signifikant positiven Einfluss auf die Preise, wobei die Pass-Through-Elastizität einen plausiblen Wert besitzt (vgl. Nijs *et al.* 2010) und Promotionaktivitäten zu signifikanten Preisreduktionen führen. Der Interaktionsterm ist zwar nur auf dem 10%-Niveau signifikant von Null verschieden, das positive Vorzeichen des Effekts ist aber sinnvoll interpretierbar. Es bedeutet, dass Händler in Wochen, in denen sie sowieso Promotionaktivitäten durchführen, einen größeren Anteil von Herstellerpreisnachlässen an die Konsumenten weitergeben. Hier steigt die Pass-Through-Elastizität von 0,397 auf $0,508 = 0,397 + 0,111$. Alles in allem lässt sich schlussfolgern, dass inhaltlich und statistisch geeignete Instrumente für die Untersuchung vorliegen.

3.5.3 Ergebnisse

Bevor auf die Ergebnisse der Schätzung eingegangen wird, ist eine kurze Diskussion von Problemen in der Modellierung bzw. Schätzung ange-

bracht, die im Rahmen der Untersuchung der Thunfischdaten auftraten.

- *Heterogenität nur im Preisparameter:* Eine Modellspezifikation, in der alle Parameter heterogen sind und miteinander korrelieren, führt mit den Thunfischdaten zu schlecht konvergierenden und unstabilen MSL-Schätzungen. Vor allem in Verbindung mit zeitvariablen Parametern zeigten vorläufige Schätzungen, dass nur Preisparameterheterogenität eine Rolle spielt. Aus diesem Grund wird nachfolgend auch nur die Preisvariable heterogen spezifiziert.²⁷
- *Modellierung der Saisonalität:* Saisonalität spielt für den Absatz von Dorsentunfisch eine Rolle. Chevalier *et al.* (2003) sowie Meza und Sudhir (2006) verwenden Indikatorvariablen und zeigen, dass bspw. die Preiselastizität in der Fastenzeit (absolut gesehen) steigt. Solche Indikatoren lieferten zusammen mit zeitvariablen Parametern im vorliegenden Beitrag keine zufriedenstellenden Ergebnisse (z. B. höhere Schätzunsicherheit), was aber wegen der flexiblen Random Walk-Prozesse auch nicht weiter verwunderlich ist.

Vielversprechender ist dagegen eine Berücksichtigung der Saisonalität in der *Varianz der UMC*. Dies ist im Rahmen der KF-Schätzung problemlos möglich, indem die Varianz als Funktion der Fastenzeitindikatoren modelliert wird:

$$\sigma_{\varepsilon jt}^2 = \sigma_{\varepsilon j}^2 \cdot \exp(\tau \cdot \mathbb{I}(t = \text{Fastenzeit})), \quad \text{für } j = 1, \dots, J. \quad (3.38)$$

Für $\tau \neq 0$ wird hierdurch (zeitlich) *bedingte Heteroskedastizität* modelliert und der KF arbeitet dann in Wochen der Fastenzeit anders als in „normalen“ Wochen. Das ist nachvollziehbar, da die Ergebnisse von Chevalier *et al.* (2003) bzw. Meza und Sudhir (2006) den Schluss zulassen, dass die Thunfischnachfrage in der Fastenzeit volatiler ist (folglich sollte $\tau > 0$ sein).

- *Komplexere Preismodelle:* Komplexere Preismodelle als (3.37) wurden ausgetestet, z. B. mit markenspezifischen Kosten-, Promotion- und Interaktionseffekten; diese lieferten aber alle lediglich einen marginal besseren Datenfit. Da der Aufwand im Rahmen einer gemeinsamen Schätzung von Nachfrage- und Preismodell stark mit der An-

²⁷ Anwendungen, die nur Preisparameterheterogenität spezifizieren bzw. in den Daten „finden“, sind in der Marketingliteratur nicht ungewöhnlich (siehe Kim 1995; Draganska und Klapper 2007; Chung *et al.* 2013).

zahl der Parameter steigt, ist eine *sparsam parameterisierte* Modellversion von Vorteil. Vorläufige Analysen (ohne Preisparameterheterogenität) zeigten, dass die verschiedenen Preismodelle keinen nennenswerten Einfluss auf die Gesamtergebnisse besitzen. Aus diesem Grund wird die einfachere Version mit Markenkonzstanten und ohne markenspezifische Effekte bevorzugt.

- *TVP-MXL (IV)-Modell*: Eine sinnvolle Schätzung des TVP-MXL (IV)-Modells basierend auf den Thunfischdaten ist nicht möglich. Die Likelihoodoptimierung konvergiert zwar, die Parameterresultate fallen aber aus inhaltlichen Gesichtspunkten derart schlecht aus, dass von einer Verwendung im Rahmen des Modellvergleichs abgesehen wird. Es ist Dynamik im Preisparameter vorhanden, dieser nimmt aber in vielen Wochen positive Werte an, was ökonomisch unplausibel ist. Zudem ist die Preisparameterheterogenität extrem hoch. Auch dieses Resultat zeigt, dass sich das TVP-MXL (IV)-Modell nicht zur Untersuchung zeitvariabler endogener Preiseffekte eignet.

Tabelle 3.5 stellt die *Schätzergebnisse der konstanten Parameter* für das TVP-MXL- und das TVP-MXL (CF)-Modell dar. Viele Parameter sind signifikant von Null verschieden und besitzen plausible Vorzeichen. Es gibt zudem Ähnlichkeiten über beide Modelle hinweg und vorhandene Unterschiede in den Werten liegen oft im Rahmen der Schätzungenauigkeit.

Insbesondere die *Dynamik- und Störtermparameter* beider Modelle sind signifikant positiv. Daher spielen UMC und Dynamik unabhängig von der Endogenitätsberücksichtigung eine wichtige Rolle in den Thunfischdaten. τ ist ebenfalls signifikant positiv und somit gibt es während der Fastenzeit extremere Nachfrageschocks. Für beide Modelle erhöht sich die UMC-Varianz fast um das vierfache (TVP-MXL (CF)-Modell: $\exp(1,301) \approx 3,673$).

Heterogenität ist nur für das TVP-MXL (CF)-Modell signifikant und zwar auf dem 10%-Niveau. Weil $\sigma_{v\beta}$ in beiden Modellen aber einen deutlich positiven Wert aufweist und wichtig zur Abbildung realistischer Substitutionsmuster ist (Dubé *et al.* 2002), wird trotzdem an der heterogenen Modellspezifikation festgehalten.

Im Fokus der empirischen Studie stehen die Ergebnisse bzgl. der *Endogenität*. Die ρ_j -Parameter für BB und CS sind signifikant von Null verschieden. Die gemeinsame Nullhypothese, dass es generell kein Endogenitätsproblem gibt, ist ebenfalls abzulehnen, wie ein *Likelihood-Quotienten-Test* mit

Tabelle 3.5: MSL-Schätzergebnisse für die untersuchten Modelle

Parameter	TVP-MXL		TVP-MXL (CF)	
	Schätzung	Std. Err.	Schätzung	Std. Err.
<i>Nutzenfkt.</i>				
$\bar{\alpha}_{BB0}$	-6,568	0,205	-6,339	0,203
$\bar{\alpha}_{CS0}$	-6,416	0,199	-6,227	0,198
$\bar{\alpha}_{HH0}$	-7,557	0,285	-7,214	0,283
$\bar{\alpha}_{SK0}$	-5,676	0,172	-5,501	0,172
$\bar{\alpha}_{Promo0}$	-0,041	0,182	0,128	0,183
$\bar{\beta}_0$	-7,034	0,963	-5,777	0,916
<i>Störterm</i>				
$\sigma_{\varepsilon BB}$	0,333	0,018	0,368	0,036
$\sigma_{\varepsilon CS}$	0,396	0,021	0,410	0,024
$\sigma_{\varepsilon HH}$	0,284	0,017	0,285	0,018
$\sigma_{\varepsilon SK}$	0,391	0,018	0,402	0,024
τ	1,363	0,173	1,301	0,170
<i>Dynamik</i>				
$\sigma_{\eta \alpha BB}$	0,078	0,018	0,085	0,018
$\sigma_{\eta \alpha CS}$	0,068	0,021	0,069	0,021
$\sigma_{\eta \alpha HH}$	0,121	0,020	0,118	0,020
$\sigma_{\eta \alpha SK}$	0,048	0,012	0,043	0,012
$\sigma_{\eta \alpha Promo}$	0,044	0,015	0,040	0,014
$\sigma_{\eta \beta}$	0,423	0,049	0,434	0,048
<i>Heterogenität</i>				
$\sigma_{\nu \beta}$	0,787	0,807	1,127	0,690
<i>Endogenität</i>				
ρ_{BB}			-0,546	0,111
ρ_{CS}			-0,252	0,106
ρ_{HH}			-0,020	0,164
ρ_{SK}			-0,208	0,118
\mathcal{L}	6.455,150		6.465,512	

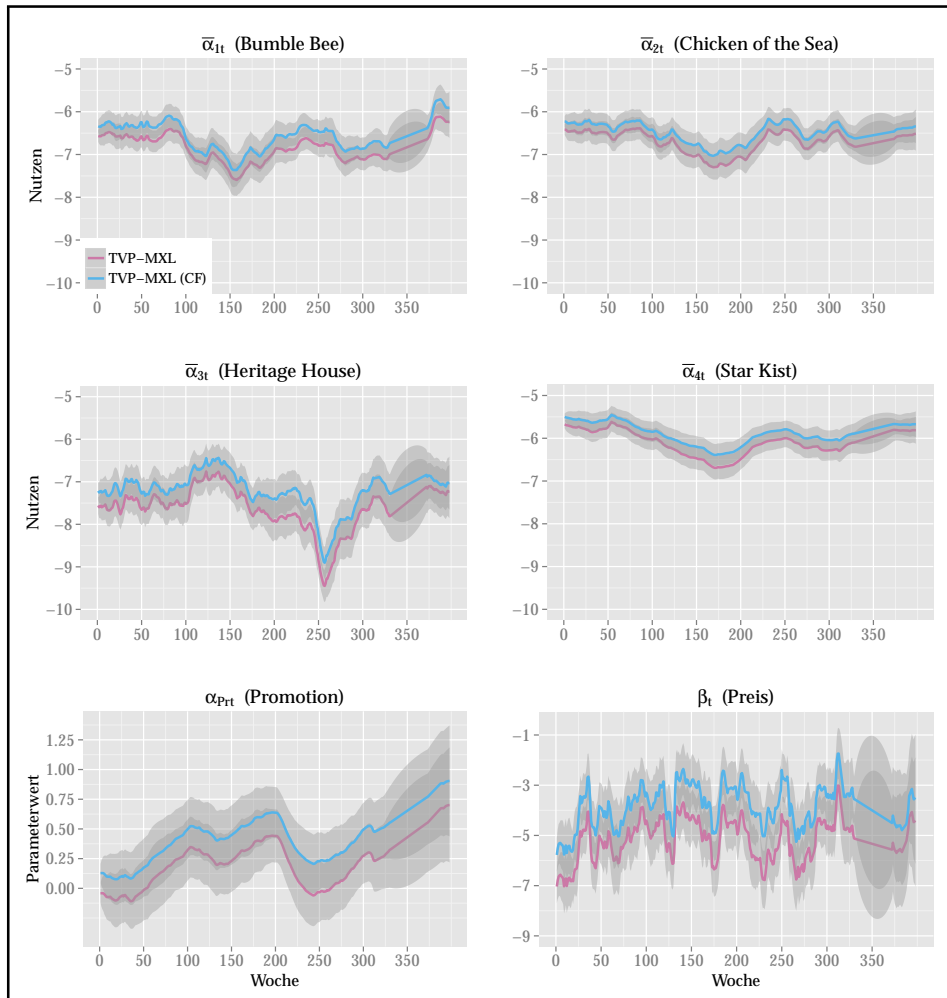
Anmerkung: Fettgedruckte Parameter sind wenigstens auf dem 5 %-Niveau signifikant von Null verschieden (zweiseitig). Der \mathcal{L} -Wert bezieht sich nur auf das Nachfragemodell und den Term der Jacobi-Determinante, damit beide Modelle vergleichbar sind. Das TVP-MXL (CF)-Modell wird allerdings einstufig inkl. Preismodell geschätzt. Die Ergebnisse des Preismodells sind aber mit denen der Preisregression (3.37) nahezu identisch, weshalb auf eine erneute Darstellung aus Platzgründen verzichtet wird.

einem hochsignifikant LR -Wert von $-2 \cdot (6.455,150 - 6.465,512) = 20,724$ bei $k = 4$ Freiheitsgraden zeigt (Greene 2011, S. 566 f.). Dieser Test zeigt ebenfalls, dass das TVP-MXL(CF)-Modell dem TVP-MXL-Modell statistisch überlegen ist, da sich beide Modelle nur bzgl. der Endogenitätsberücksichtigung unterscheiden. Erwähnenswert ist das negative Vorzeichen der ρ_j -Parameter. Oftmals wird von einer positiven Korrelation zwischen Preisen und UMC ausgegangen (Villas-Boas und Winer 1999), negative Korrelationen sind aber natürlich auch möglich (Park und Gupta 2009) und deuten darauf hin, dass es unbeobachtete Absatzförderungsmaßnahmen gibt (z. B. Herstellercoupons), die mit Preisreduktionen einhergehen (Train 2009).

Der oberste Block in Tabelle 3.5 bezieht sich auf die *Startwerte der Nutzenparameter* zum Zeitpunkt $t = 0$. Bevor näher auf die Verläufe dieser zeitvariablen Parameter eingegangen wird, zeigen bereits diese Startwerte eine Endogenitätsverzerrung. Die Parameter des TVP-MXL-Modells fallen wegen der zuvor erwähnten negativen Korrelation zu niedrig aus. Vor allem die Preissensitivität ist ohne Endogenitätsberücksichtigung damit (absolut gesehen) zu hoch. Abbildung 3.3 zeigt die *Verläufe der zeitvariablen Parameter* für beide Modelle. Sämtliche Nutzenparameter des TVP-MXL-Modells besitzen kleinere Werte verglichen mit dem TVP-MXL(CF)-Modell. Zum einen passen diese Verzerrungen zu denen der Startwerte und zum anderen ist abermals die bereits im Rahmen der Simulationsstudie diskutierte Parallelverschiebung der Parameter zu erkennen. Die Abbildung zeigt indes deutlich, dass alle Parameter zeitlich variieren. Diese Variation ist zudem von Parameter zu Parameter sowohl in ihrer Form als auch ihrer Intensität unterschiedlich und entspricht keinesfalls monotonen Trends. Hier zeigt sich die Überlegenheit der Random Walk-Spezifikation, weil hierdurch a priori keine Verlaufsannahmen notwendig sind (Lachaab *et al.* 2006).

Die Abbildung zeigt ebenfalls die 95%-*Konfidenzintervalle* der Parameter. Zwei Punkte sind auffällig: 1.) Die Konfidenzintervalle überlappen sich für beide Modelle zu jedem Zeitpunkt. Auf diese Weise lässt sich aber nicht überprüfen, ob die Endogenitätsverzerrung signifikant ist. Ein solcher Test wird – wie zuvor in Abschnitt 3.2.3 erwähnt – anhand der ρ_j -Parameter vollzogen. 2.) In den Wochen 331 bis 372 liegen keine Daten vor. Der KF und der KS können mit diesen fehlenden Werten aber problemlos umgehen und ermitteln trotzdem Schätzwerte. Die Konfidenzintervalle weisen für diesen Zeitraum aber charakteristische Vergrößerungen auf, weil die Schätzungen entsprechend unsicherer sind.

Abbildung 3.3: Grafische Darstellung der zeitlichen Parameterverläufe



Anmerkung: Die 95 %-Konfidenzintervalle sind dunkelgrau hinterlegt eingezeichnet.

Die *Markennutzen* für BB, CS und SK besitzen ähnliche Entwicklungen, mit fallenden Verläufen in den ersten 150 Wochen und einem anschließenden Anstieg. Absolut gesehen hat SK die höchsten Werte, BB und CS liegen auf einem vergleichbaren Niveau. HH hat oftmals die niedrigsten Werte (mit Ausnahme der Wochen 100 bis 150), was aber nicht weiter verwunderlich ist, weil HH die Handelsmarke von DFF ist. Ferner besitzt HH einen abweichenden Verlauf im Vergleich zu BB, CS und SK. Nach einer relativ stabilen Entwicklung in den ersten 240 Wochen gibt es einen Einbruch, wovon sich die Marke aber innerhalb eines Jahres erholt.

Der *Promotionparameter* weist grundsätzlich einen steigenden Trend über den gesamten Zeitraum der Daten auf. Es gibt aber zwei Rückgänge, nach 100 und 200 Wochen. Dieses Ergebnis belegt, dass der Effekt von Promotionaktivitäten in 1990er Jahren bei DFF zugenommen hat.

Der *Preisparameter* ist sehr starken Schwankungen unterworfen, was auch schon aus dem hohen σ_{η_β} -Wert von ca. 0,4 hervorgeht. Für das TVP-MXL (CF)-Modell liegen die Preisparameter zwischen $-5,777$ und $-1,740$, mit einem Mittelwert von $-3,853$. Der Preisparameter des TVP-MXL-Modells hat dagegen einen Mittelwert von $-5,177$ und damit beträgt die Endogenitätsverzerrung im Durchschnitt $-1,324$.

Wie bereits angemerkt, verwenden einige Autoren einen Interaktionsterm für den *Fastenzeitindikator* und den Preis, um saisonale Variation im Preiseffekt zu modellieren. Die Korrelation des Fastenzeitindikators mit dem Preisparameter des TVP-MXL (CF)-Modells beträgt $-0,176$. Damit lässt sich das Ergebnis von Chevalier *et al.* (2003) replizieren, die während der Fastenzeit eine (absolut gesehen) höhere Preissensitivität ermitteln. Eine *GLS-Dummyregression* des Preisparameters (basierend auf dem TVP-MXL (CF)-Modell inkl. Berücksichtigung der Schätzunsicherheit) auf den Fastenzeitindikator zeigt, dass die Fastenzeit einen hochsignifikant negativen Effekt von $-0,644$ besitzt. Aufgrund der sehr niedrigen Varianzaufklärung von gerade mal 3,7 % ist allerdings evident, dass nur ein Bruchteil der zeitlichen Variation auf den Fastenzeitindikator zurückgeht. Hierdurch wird abermals deutlich, welchen Nutzen eine flexible Spezifikation der zeitvariablen Parameter besitzt.

Es sei daran erinnert, dass es sich bei $\bar{\beta}_t$ nur um den Mittelwert des normalverteilten Preisparameters handelt. Für das TVP-MXL (CF)-Modell impliziert der Wert von 1,127 für σ_{v_β} , dass 90 % der Populationswerte zwischen $-1,851$ und $1,851$ liegen, wenn der Mittelwert auf Null gesetzt wird. Hieraus ist zu erkennen, dass Variationen im Preisparameter aufgrund von Heterogenität und Dynamik durchaus ähnliche Größenordnungen besitzen, obwohl es sich um gänzlich unterschiedliche Phänomene handelt.

Im *Ganzen* zeigt die Modellschätzung für die Thunfischkategorie von DFF, dass Heterogenität und Endogenität relevant sind und sämtliche Parameter über die Zeit variieren.

3.5.4 Implikationen

Zur Untersuchung, welche *praktischen Implikationen* aus den zuvor gewonnenen Ergebnissen und der Verwendung des „richtigen“ Modells folgen, werden Preisoptimierungen durchgeführt und miteinander verglichen, die auf dem TVP-MXL (CF)- und TVP-MXL-Modell fußen.²⁸

Hierzu sind zunächst mehrere *Annahmen* notwendig:

1. Der Händler maximiert Woche für Woche den Deckungsbeitrag für die Thunfischkategorie in allen Geschäften.
2. Konkurrierende Händler werden ignoriert.
3. Der Händler optimiert nur die Preise. Werte für die Promotion- und Kostenvariablen sind dem Händler zwar bekannt, sie werden aber als gegeben betrachtet.
4. Die Grundlage für die Vergleiche bilden die ersten 300 Wochen²⁹ aus dem Datensatz.³⁰
5. Die tatsächlichen Preise dienen als Grundlage für eine Status Quo-Betrachtung.

Für das *Optimierungsproblem* des Händlers gilt

$$\pi_t = \max_{\{p_{jt}\}_{j=1}^J} M_t \cdot \sum_{j=1}^J (p_{jt} - c_{jt}) \cdot s_{jt}(\mathbf{p}_t), \quad \forall t. \quad (3.39)$$

Innerhalb einer Woche sind die Terme des Optimierungsproblems aufgrund der Struktur des Logitmodells miteinander verbunden, der Preis *einer* Marke besitzt hier einen Effekt auf *sämtliche* Marken. Für alle Zeitpunkte lässt sich (3.39) allerdings einzeln maximieren, weil die Preise im

²⁸ Die normative Modellierung orientiert sich u. a. an Kim *et al.* (1995) sowie Vilcassim und Chintagunta (1996).

²⁹ Wegen vieler fehlender Werte am Ende des Untersuchungszeitraums werden nur 300 der insg. 337 Wochen berücksichtigt.

³⁰ Diese Annahme ist notwendig, damit eine Optimierung über viele Wochen (hier 300) mit möglichst realistischen Kosten- und Promotionwerten stattfinden kann. Es sei in Erinnerung gerufen, dass alle Nutzenparameter zeitvariabel sind und ein einperiodiges Vorgehen mit Durchschnittswerten für die Kosten- und Promotionvariable (= „Durchschnittswoche“) nicht die gesamte Bandbreite möglicher Konsumentenreaktionen abbilden würde.

Modell *keine intertemporalen Effekte* besitzen. Die *Bedingung erster Ordnung* für ein Optimum von (3.39) lautet:

$$s_{jt}(\mathbf{p}_{jt}^*) + \sum_{k=1}^J \frac{\partial s_{kt}(\mathbf{p}_t^*)}{\partial p_{jt}} \cdot (p_{kt}^* - c_{kt}) \equiv 0, \text{ mit } j = 1, \dots, J. \quad (3.40)$$

Gleichung (3.40) lässt sich kompakt als Matrixgleichung schreiben:

$$\mathbf{p}_t^* = \mathbf{c}_t - \Delta_t(\mathbf{p}_t^*)^{-1} \cdot \mathbf{s}_t(\mathbf{p}_t^*). \quad (3.41)$$

Der Vektor \mathbf{p}_t^* enthält die optimalen Preise für alle J Marken zum Zeitpunkt t . Die Matrix $\Delta_t(\mathbf{p}_t^*)$ fasst sämtliche ersten Ableitungen der Marktanteile nach dem Vektor der Preise zusammen (vgl. z. B. Chintagunta *et al.* 2003). $\mathbf{s}_t(\mathbf{p}_t^*)$ ist der Vektor der Marktanteile zum Zeitpunkt t , evaluiert an der Stelle der optimalen Preise \mathbf{p}_t^* und die J Werte der variablen Kosten sind in \mathbf{c}_t zu finden.

Weil $\Delta_t(\mathbf{p}_t^*)$ und $\mathbf{s}_t(\mathbf{p}_t^*)$ vom optimalen Preis abhängen, liefert (3.41) nicht die mit der Optimierung „gesuchten“ optimalen Preise, sondern lediglich eine *Optimalitätsbedingung* für deren Höhe. Diese Gleichung lässt sich aber als *kontrahierende Abbildung* iterativ ausführen, um eine numerische Lösung für die optimalen Preis zu erhalten (Morrow und Skerlos 2011). Der Algorithmus ist robust, einfach anzuwenden und konvergiert zügig.

Ausgehen von einem Startvektor $\mathbf{p}_t^{(0)}$ (z. B. den Vektor der Durchschnittspreise) wird

$$\mathbf{p}_t^{(h+1)} = \mathbf{c}_t - \Delta_t(\mathbf{p}_t^{(h)})^{-1} \cdot \mathbf{s}_t(\mathbf{p}_t^{(h)}) \quad (3.42)$$

solange iterativ berechnet, bis $\|\mathbf{p}_t^{(h+1)} - \mathbf{p}_t^{(h)}\| < 1 \cdot 10^{-14}$ gilt. Als Ergebnis folgt der Vektor der optimalen Preise \mathbf{p}_t^* und diese Optimierung ist für alle $t = 1, \dots, T$ zu wiederholen.

Für die Untersuchung der Konsequenzen einer Modellierung mit und ohne Endogenität erfolgen Optimierungen mit dem TVP-MXL (CF)- und dem TVP-MXL-Modell. Die optimalen Preise und die tatsächlichen Preise werden anschließend in das TVP-MXL (CF)-Modell (das „wahre“ Modell) zur Berechnung der (fiktiven) Marktanteile und Deckungsbeiträge eingesetzt (vgl. das analoge Vorgehen bei Dubé *et al.* 2008, S. 423 ff.). Hierdurch

wird die Vergleichbarkeit der Ergebnisse gewährleistet, da sich die *Evaluation* der unterschiedlichen Preissetzungen auf *dasselbe Modell* bezieht.

Tabelle 3.6: Modellvergleich auf Basis der Preisoptimierung

Tatsächliche Preise							
Marke	Preis		Marktanteil (kond.)**		Marktanteil (unkond.)**		Deckungs- beitragsrate
BB	0,764	(0,108)*	22,055	(12,887)	0,529	(0,488)	28,356 (10,322)
CS	0,792	(0,086)	22,060	(12,743)	0,557	(1,063)	26,977 (9,016)
HH	0,687	(0,067)	19,483	(11,603)	0,467	(0,625)	27,096 (9,911)
SK	0,795	(0,096)	36,401	(15,716)	0,929	(1,204)	28,484 (8,968)
Händlerdeckungsbeitrag			8.282,926 (4.977,618)				
Optimale Preise des TVP-MXL (CF)-Modells							
Marke	Preis		Marktanteil (kond.)		Marktanteil (unkond.)		Deckungs- beitragsrate
BB	0,727	(0,086)	23,059	(8,718)	0,695	(0,528)	25,497 (4,981)
CS	0,775	(0,069)	19,948	(7,183)	0,618	(1,017)	25,932 (5,154)
HH	0,661	(0,063)	19,722	(10,285)	0,641	(1,035)	24,882 (4,490)
SK	0,761	(0,077)	37,271	(9,255)	1,138	(1,555)	25,800 (5,152)
Händlerdeckungsbeitrag			9.820,554 (5.556,297)				
Optimale Preise des TVP-MXL-Modells							
Marke	Preis		Marktanteil (kond.)		Marktanteil (unkond.)		Deckungs- beitragsrate
BB	0,673	(0,067)	22,945	(8,385)	0,880	(0,589)	19,804 (2,727)
CS	0,714	(0,053)	20,188	(7,058)	0,789	(1,096)	19,869 (2,729)
HH	0,618	(0,053)	19,275	(10,015)	0,792	(1,139)	19,726 (2,645)
SK	0,702	(0,056)	37,592	(9,008)	1,447	(1,590)	19,852 (2,821)
Händlerdeckungsbeitrag			9.519,483 (5.594,615)				

Anmerkung: (*) Mittelwert und Standardabweichung in Klammern, berechnet über wöchentliche Werte. Die geringen Abweichungen verglichen mit den Werten in Tabelle 3.4 sind durch den kleineren Zeitraum der Daten (300 Wochen) zu erklären. (**) Die Berechnung sämtlicher Marktanteile basiert auf dem TVP-MXL (CF)-Modell. Marktanteile und Deckungsbeitragsraten $((\text{Preis} - \text{Kosten}) / \text{Preis})$ werden in % angegeben. Die konditionalen Marktanteile beziehen sich nur auf die Absätze innerhalb der Kategorie. Dagegen schließen die unkonditionalen Marktanteile die Nichtwahloption mit ein. BB hat folglich tatsächliche Marktanteile im Durchschnitt von ca. 22 % (innerhalb der Kategorie) bzw. ca. 0,5 % (inkl. Nichtwahloption). Die Preise und Händlerdeckungsbeiträge (jeweils in \$) beziehen sich auf 6 oz-Dosen.

Tabelle 3.6 bietet einen Überblick zum Modellvergleich auf Basis der Preisoptimierungen. Der obere Teil bezieht sich zunächst auf die *tatsächlichen Preise* und dient als Status Quo. Die vom TVP-MXL(CF)-Modell mit den *tatsächlichen* Preisen generierten Marktanteile der Marken (Kategoriekauf vorausgesetzt) belegen, dass das TVP-MXL (CF)-Modell auf Marktanteils-

ebene zufriedenstellende Prognosen liefert, da die Differenz zu den wahren Marktanteilen gering ausfällt. Für letztere gelten die Durchschnittswerte (vgl. Tabelle 3.4) 22,3 %, 22,9 %, 19,6 % und 35,2 % (jeweils für BB, CS, HH und SK). Die tatsächlichen Deckungsbeitragsraten liegen zwischen 27 % und 28,5 % und der Gesamtdeckungsbeitrag des Händlers für die Thunfischkategorie würde bei den tatsächlichen Preisen modellgemäß im Durchschnitt ca. 8.283 \$ pro Woche betragen.

Der *mittlere Teil* der Tabelle stellt die Ergebnisse der Preisoptimierung auf Basis des TVP-MXL(CF)-Modells dar. Die optimalen Preise sind (im Durchschnitt) etwas geringer als die tatsächlichen Preise, mit Differenzen zwischen 0,017 \$ (CS) und 0,037 \$ (BB). Ähnliches gilt auch für die Deckungsbeitragsraten. Die geringeren Preise im Optimum haben zur Folge, dass die Attraktivität der Nichtwahloption sinkt. Zudem ändern sich auch die auf den Kategoriekauf bedingten Marktanteile. CS verliert im Optimum ca. zwei Prozentpunkte Marktanteil und liegt jetzt auf dem Niveau von HH und BB, der ehemalige direkte Konkurrent von CS, legt einen Prozentpunkt zu. SK kann seine Marktführerschaft ausbauen (ca. 0,9 Prozentpunkte) und bleibt damit die beliebteste Thunfischmarke bei DFF. Der Händlerdeckungsbeitrag beträgt bei Verwendung optimaler Preise etwas mehr als 9.820,5 \$ pro Woche. Ein Deckungsbeitragspotenzial von knapp 1.538 \$ wird durch den Einsatz der *tatsächlichen* Preise somit pro Woche nicht ausgeschöpft. Hochgerechnet auf ein Jahr ergibt sich ein mittels modellbasierter Preisoptimierung erschließbarer Deckungsbeitrag in der Thunfischkategorie von rund 80.000 \$ (exakt: 79.976 \$). So viel dürfte folglich die modellgestützte Preisoptimierung (Datenbeschaffung und -auswertung sowie Umsetzung) pro Jahr maximal kosten, damit sie für den Händler rentabel wäre.

Der *untere Teil* der Tabelle illustriert, was passiert, wenn beim TVP-MXL-Modell Endogenität ignoriert wird. Die vermeintlich optimalen Preise fallen viel geringer aus, z. B. fast $-0,1$ \$ für BB und SK im Vergleich zu den *tatsächlichen* Preisen. Aufgrund der niedrigeren Preise sind auch die Deckungsbeitragsraten kleiner und liegen jetzt alle unter 20 %. Dies ist die Konsequenz der Endogenitätsverzerrung, weil hierdurch die Preissensitivität fälschlicherweise (absolut gesehen) überschätzt wird. Durch die hohen Preisreduktionen steigen die Thunfischmarktanteile im Vergleich zur Nichtwahloption sehr stark an. Der Händlerdeckungsbeitrag liegt mit ca. 9.519 \$ zwischen dem Status Quo und dem Resultat des TVP-MXL(CF)-Modells. Wird wieder ein ganzes Jahr betrachtet, wäre die Endogenitätsberücksichtigung 15.652 \$ wert.

Die normative Analyse zeigt, dass die optimalen Preise des TVP-MXL-Modells sowie die tatsächlichen Preise – gemessen an den aus einem leistungsfähigeren Nachfragemodell ableitbaren optimalen Preisen – suboptimal sind, und dass es deutliche Deckungsbeitragsunterschiede im Vergleich zum Optimum gibt. Das Ignorieren der Preisendogenität kostet den Händler einen fünfstelligen Betrag pro Jahr. Dieser wäre also gut beraten, das TVP-MXL (CF)-Modell zur Preisoptimierung einzusetzen.

3.6 Zusammenfassende Schlussbetrachtung

Der vorliegende Beitrag behandelt Logitmodelle für aggregierte Absatzdaten, die neben Konsumentenheterogenität und Preisendogenität zusätzlich die zeitliche Variation sämtlicher Parameter erlauben. Gerade im Marketing ist die Beachtung solcher Dynamik wichtig (Leeftang *et al.* 2009). Die Sinnhaftigkeit einer Unterstellung zeitlich konstanter Nutzenfunktionen der Konsumenten ist deshalb fragwürdig (Lachaab *et al.* 2006). Da zudem die in Marktdaten antreffbaren Preise nachweislich als endogene Größen zu verstehen sind (Villas-Boas und Winer 1999), erscheint es bei der Nachfragemodellierung angebracht zu sein, beide Phänomene (Dynamik und Endogenität) explizit in einer Modellspezifikation zu berücksichtigen. Ist nämlich der Modellparameter einer endogenen Variablen zeitvariabel, führen Standardmodelle bzw. Schätzverfahren (2SLS, GMM, ...) zur Behandlung von Endogenitätsverzerrungen nicht zu einem validen Ergebnis. Als geeignete Lösung dieses Problems wird hier ein CF-Ansatz vorgestellt und untersucht.

Sowohl mittels Simulationen als auch an anhand einer empirischen Studie werden zahlreiche *Erkenntnisse* gewonnen.

- Der postulierte CF-Ansatz ist in der Lage, zeitvariable Parameter für aggregierte Logitmodelle, unter Berücksichtigung von Heterogenität und Endogenität, konsistent zu bestimmen.
- U. U. kann es dazu kommen, dass Verzerrungen ohne Endogenitätskorrektur kleiner ausfallen als bei einer falschen Korrektur.
- Eine falsche Endogenitätskorrektur führt dazu, dass dynamische Effekte sämtlicher Modellvariablen verzerrt sein können. Diese Verzerrungen betreffen nicht nur die absoluten Parameterwerte zu einem gegebenen Zeitpunkt, sondern vor allem die zeitlichen Verläufe.

- Die empirische Studie anhand der Thunfischdaten des Händlers DFF belegt empirisch die Relevanz eines Logitmodells mit Heterogenität, Endogenität und Dynamik.
- Das Deckungsbeitragspotenzial einer modellbasierten Preisoptimierung beträgt für DFF rund 80.000 \$ pro Jahr.
- Wird Endogenität bei der modellbasierten Preisoptimierung ignoriert, führen Verzerrungen in den geschätzten Preissensitivitäten zu suboptimalen Preisen. DFF entgeht hierdurch ein Deckungsbeitrag von mehr 15.000 \$ pro Jahr.

Das hier postulierte Vorgehen hat natürlich auch Schwächen und es sind mehrere *Erweiterungen* denkbar bzw. lohnenswert für zukünftige Arbeiten:

- Auch für das Preismodell ist eine Spezifikation mit zeitvariierenden Parametern denkbar, um Kosteneffekte flexibler abzubilden (Kim und Kim 2011).
- Das vorgestellte Modell behandelt lediglich die Markenwahl, die Kaufmenge pro Kauf wird ignoriert. Für Kategorien, deren Produkte gut lagerfähig sind, wäre eine entsprechende Erweiterung sinnvoll (Nair *et al.* 2005).
- Die MSL-Schätzung ist nicht gut geeignet für umfangreiche Modelle mit sehr vielen Marken und erklärenden Variablen. In solchen Fällen wäre ein Ansatz basierend auf der bayes'schen Schätztheorie vielversprechend (Jiang *et al.* 2009; Petris *et al.* 2009).
- Alle zeitvariablen Parameter wurden als Random Walk-Prozesse modelliert, allerdings sind auch alternative dynamische Modelle denkbar. Mit fundierten Vorkenntnissen und entsprechenden Daten könnten über erklärende Variablen in den Zustandsgleichungen langfristige Marketing-Mix-Effekte modelliert werden (vgl. hierzu auch Abschnitt 2.2 im vorherigen Kapitel).

Zusammenfassend betrachtet steht mit dem vorgestellten CF-Ansatz ein leistungsfähiges und sinnvolles Werkzeug zur empirischen Untersuchung dynamischer Märkte zur Verfügung.

Kapitel 4

Die Messung zeitvariabler Markenwerte mit Haushaltspaneldaten

„[T]he dynamic nature of brand equity and of the behavioral processes that underlie the formation and evolution of brand equity, suggest that it is important for brand equity measurement models to capture dynamics.“

Erdem *et al.* (1999, S. 312)

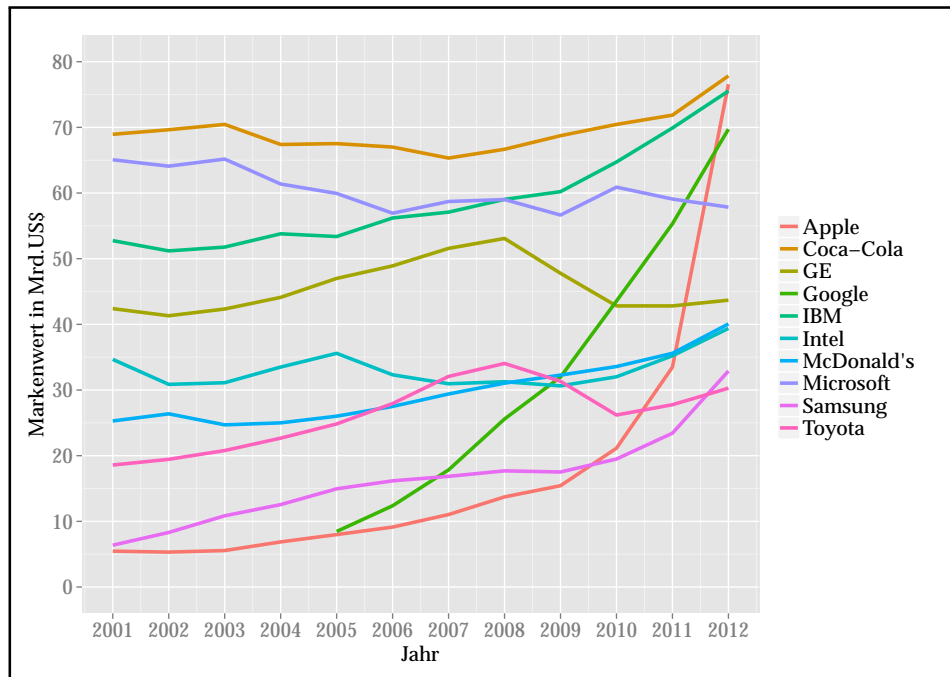
4.1 Einleitung

Marken haben heutzutage eine große wirtschaftliche Bedeutung und stellen für Unternehmen wertvolle immaterielle Vermögensgegenstände dar (Ailawadi *et al.* 2003; Keller und Lehmann 2006). Abbildung 4.1 zeigt z. B. *Markenwerte* und deren zeitliche Entwicklung für die zehn Marken mit den weltweit höchsten Werten im Jahr 2012. Zwei Punkte sind deutlich zu erkennen und lassen sich als „stylized facts“ interpretieren:

- Internationale „Top-Marken“ besitzen Markenwerte in Milliardenhöhe und
- Markenwerte variieren z. T. erheblich über die Zeit.

Konsumgüterhersteller tätigen hohe *Marketinginvestitionen*, um ihre Marken aufzubauen, zu stärken sowie vor dem Verfall zu schützen und da-

Abbildung 4.1: Zeitliche Entwicklung der Markenwerte internationaler Top-Marken



Quelle: Eigene Darstellung in Anlehnung an <http://www.interbrand.com/de/best-global-brands/BGB-Interactive-Charts.aspx> (letzter Abruf: 21 Juni 2012).

durch langfristig Erträge zu sichern.¹ Procter & Gamble gab bspw. 2012 weltweit allein für Werbung über 9 Mrd. \$ aus (Johnson 2012) und Pepsi-Co erhöhte im selben Jahr die Ausgaben um 500 Mio. \$ auf über 2 Mrd. \$ (Zmuda 2012). Damit solche Budgetentscheidungen nicht im „Blindflug“ getroffen werden, müssen Unternehmen den Markenwert möglichst genau kennen, was unmittelbar zur Frage führt, wie der Wert einer Marke überhaupt gemessen werden kann. So einfach, wie es Abbildung 4.1 mit ihren Zahlen vermuten lässt, ist die Markenwertmessung nicht (Farris *et al.* 2010, S. 136 f.). Aus diesem Grund ist dieses Thema bis heute für die Marketingforschung und -praxis gleichermaßen von hohem Interesse (Ailawadi *et al.* 2003; Deighton *et al.* 2012).

Der Wert einer Marke wird als der zusätzliche Wert definiert, den ein Produkt erhält, wenn es den Markennamen trägt und allein schon deshalb leicht innerhalb einer Produktkategorie identifiziert werden kann (Farquhar 1989; Aaker 1991; Ailawadi *et al.* 2003; Sriram *et al.* 2007; Hernández-

¹ Dies gilt ebenso für B2B-Unternehmen (Keller und Kotler 2012).

Mireles und Fok 2011).² Zur Markenwertmessung wurden in der Literatur bereits eine große Zahl unterschiedlicher Metriken vorgeschlagen und hinsichtlich ihrer Eignung evaluiert (vgl. hierzu z. B. Agarwal und Rao 1996 für konsumentenbasierte Maße). Die verschiedenen Metriken lassen sich gut entlang *zweier Dimensionen* systematisieren (vgl. z. B. Hernández-Mireles und Fok 2011). Zum einen danach, ob sie „*unverfälscht*“ sind, im Sinne einer monokausalen Rückführung auf den Effekt der Marke und einer guten Trennung von Marketing-Mix-Effekten und anderen Einflussgrößen (Keller und Lehmann 2006, S. 751). Zum anderen ist es wichtig, wie eng der *Bezug zum äußeren Verhalten* der Konsumenten ausfällt. Schließlich existiert der Markenwert nicht nur in den Köpfen der Konsumenten, sondern besitzt auch Kaufverhaltensrelevanz (Keller und Lehmann 2006). Einige Beispiele zu verschiedenen Markenwertmaßen und deren Einordnung in die zuvor genannte Systematisierung sind Tabelle 4.1 zu entnehmen.³

Tabelle 4.1: Systematisierung alternativer Maße des Markenwerts

Ist das Maß unverfälscht?	Besteht ein direkter Bezug zum äußeren Verhalten?	
	ja	nein
ja	<i>Nutzenbasierte Größen:</i> Markenkonstanten, Equalization Price, ...	<i>Psychische Größen:</i> Markenbekanntheit, Markensympathie, ...
nein	<i>Marktgrößen:</i> Marktanteil, Umsatzpremium, ...	—

Richtig gewählte *psychische Größen* (sog. „mindset metrics“) sind tendenziell unverfälscht und passen zu einem verhaltenswissenschaftlichen Markenverständnis. Sie stellen allerdings nur eine Vorstufe des Kaufverhaltens dar (Steffenhagen 2000). Zudem werden sie mittels Befragungen erhoben (Keller 1993) und sind daher möglicherweise fehlerbehaftet (Sriram *et al.* 2007, S. 61).

² Hierzu existieren alternative Definitionen, z. B. solche, die auch Logos oder Verpackungsbesonderheiten der Marke einbeziehen (Keller und Lehmann 2006). Für den vorliegenden Beitrag sind diese Unterscheidungen aber nicht relevant und der vereinfachende Fokus auf den Markennamen ist zweckmäßig. Daher findet auch eine synonyme Verwendung der Termini „Marke“ und „Markenname“ statt.

³ Alternative Systematisierungen sowie State of the Art-Auflistungen gängiger Markenwertmaße sind bei Sattler (1995, 2005) oder Keller und Lehmann (2006) zu finden.

Marktgrößen, wie etwa das Umsatzpremium, haben einen engen Bezug zum individuellen Kaufverhalten, denn sie sind schließlich die Konsequenz hieraus. Sie korrespondieren gut mit einer rein ökonomischen Markenbetrachtung, und finanzielle Aspekte der Marke lassen sich leicht ableiten. Dennoch sind sie ein durch unterschiedliche Ursachen (z. B. Marketing-Mix-Effekte oder Saisonalität) bedingtes Amalgam, und eine zeitliche Variation wird durch eine Vielzahl von Marktdynamiken beeinflusst.

Sehr günstig in beiden genannten Dimensionen schneiden *nutzenbasierte Größen* ab. Bereits zu Beginn der 1990er Jahre schlugen Kamakura und Russell vor, den Markenwert über den *Markennutzen* mithilfe diskreter Wahlmodelle basierend auf *Marktdaten* zu messen (vgl. Kamakura und Russell 1993 und auch die Konzeptualisierung von Srinivasan 1979). Da diese Modelle auf der ökonomischen Theorie der Nutzenmaximierung beruhen, ist ein enger Bezug zum Kaufverhalten sichergestellt, denn der Nutzen bestimmt die Wahl einer Marke unmittelbar (Chintagunta und Nair 2011). Gleichzeitig erlauben diese Modelle eine saubere Trennung von Effekten der Marke, des Marketing-Mix und von anderen exogenen Einflüssen (Keller und Lehmann 2006). Nutzenbasierte Maße sind demnach unverfälscht. Auch wenn nutzenbasierte Größen lediglich auf der Konsumentenebene existieren, lassen sie sich auf Marktebene hochrechnen und entsprechend interpretieren, was ebenfalls vorteilhaft ist (Sriram *et al.* 2007).

Die einleitenden Ausführungen legen nahe, dass es sich beim Markenwert um ein *dynamisches Konstrukt* handelt, welches folglich auch *zeitvariabel gemessen* werden sollte (vgl. hierzu ausführlich Hanssens und Dekimpe 2008 sowie Leeftang *et al.* 2009). Ein Problem der bekannten nutzenbasierten Maße ist jedoch, dass es sich entweder um Momentaufnahmen handelt (Querschnittsdaten) oder um Durchschnittswerte (Paneldaten ohne Dynamik), die über einen gewissen Zeitraum gebildet werden (z. B. 52 Wochen bei Kamakura und Russell 1993).⁴ Das widerspricht der dynamischen Natur des Markenwertkonstrukts und auch dem Verständnis in der Praxis. Erdem und Swait (2010, S. 221) stellen mit Bezug zu diskreten Wahlmodellen fest: „*Brand equity formation is inherently a dynamic phenomenon. Yet most of the modeling literature that directly addresses issues of brand equity measurement has taken a static view [...]*“. Hierin mag der Grund liegen, warum aktuell für Trackingzwecke des Markenwerts eher auf Marktgrößen (sind

⁴ Neben Marktdaten sind auch Daten aus Wahlexperimenten sehr beliebt, um nutzenbasierte Markenwerte zu messen (Swait *et al.* 1993; Sattler 2005). Hier stehen dann ebenfalls nur Momentaufnahmen im Fokus.

in Unternehmen auf zeitlich disaggregierter Ebene vorhanden) oder psychische Größen (lassen sich einfach regelmäßig erheben) zurückgegriffen wird.

Aktuelle Arbeiten schlagen deshalb Methoden vor, die explizit eine dynamische Messung nutzenbasierter Markenwerte ermöglichen. Diese beziehen sich aber bis auf Ausnahmen⁵ auf aggregierte Daten (vgl. Sriram *et al.* 2006, 2007 und Hernández-Mireles und Fok 2011, sowie die Kapitel 2 und 3 der vorliegenden Arbeit). *Aggregierte* Daten haben neben dem Vorteil einer besseren Verfügbarkeit allerdings auch *Nachteile*:

1. *Konsumentenheterogenität*, die im Marketing von großer Relevanz ist (Allenby und Rossi 1999), lässt sich nur eingeschränkt und mit hohem Aufwand abbilden (Chintagunta und Dubé 2005, S. 368).
2. *Kauffeedbackeffekte*⁶ sind mit aggregierten Daten ebenfalls nur schwierig zu berücksichtigen (Horsky *et al.* 2012). Eine Trennung von Marken- und Kauffeedbackeffekten ist aber für eine unverfälschte Markenwertmessung in diskreten Wahlmodellen von hoher Wichtigkeit (Keller und Lehmann 2006, S. 751).
3. Aggregierte Daten sind anfällig für Probleme aufgrund des *Aggregationsbias* (Christen *et al.* 1997). Folglich erscheint es angebracht, vorhandene Informationen auf einem möglichst disaggregierten Niveau zu nutzen, sofern dies möglich ist (Villas-Boas und Winer 1999; Chintagunta und Dubé 2005).

Für *Individualdaten* (Haushaltspaneldaten) schlagen Lachaab *et al.* (2006) und Rutz und Sonnier (2011) einen bayes'schen Ansatz vor, um (u. a.) zeitvariable Markennutzen mit Probitmodellen zu erfassen. Der hohe Implementierungs- und Schätzaufwand sowie die äußerst geringe Standardisier- und Skalierbarkeit (z. B. auf Kategorien mit vielen Marken) verhindern allerdings – zumindest momentan – deren Anwendung und Verbreitung. Vor allem in der Praxis ist niemand bereit, wochen- oder monatelang auf Ergebnisse zu warten (Sattler 2005). Des Weiteren ist keine Standardsoftware zur Anwendung des Ansatzes von Lachaab *et al.* (2006) vorhanden.

⁵ Eine ausführliche Literaturdiskussion ist in Abschnitt 4.2 zu finden.

⁶ Als Kauffeedbackeffekte werden nachfolgend generell Einflüsse vergangener Wahlentscheidungen auf die aktuelle Wahlentscheidung bezeichnet (Seetharaman 2004).

Gegenstand des vorliegenden Beitrags ist ein Vorgehen zur nutzenbasierten Messung zeitvariabler Markenwerte mit Haushaltspaneldaten, welches die beschriebenen Defizite nicht aufweist. Es ermöglicht, *State of the Art-Wahlmodelle* – die Konsumentenheterogenität, Preisendogenität, unbeobachtete Markencharakteristika (UMC) und Kauffeedbackeffekte berücksichtigen – mit einer flexiblen Spezifikation dynamischer Effekte (hier z. B. Markennutzen) zu kombinieren. Es wird eine *zweistufige Modellschätzung* verwendet, die keinen nennenswerten Zeitaufwand erfordert und sich mit bewährter Software durchführen lässt.⁷ Der hier präsentierte Ansatz ist somit *anwendungsorientiert* und *praxistauglich*. Aufbauend auf jüngeren methodischen Entwicklungen (Berry *et al.* 2004; Chintagunta *et al.* 2005) werden auf einer ersten Schätzstufe sog. *fixe Effekte (FE)* für jede Marke und jeden Zeitpunkt geschätzt (Berry 1994, S. 259). Diese FE beinhalten sämtliche Variation in den Daten, die nicht personenspezifisch ist, wozu auch die dynamischen Markennutzen gehören. Auf der zweiten Schätzstufe, mit den FE als abhängiger Variable, ist es dann mithilfe des *Kalman Filters (KF)* möglich, dynamische Markennutzen flexibel zu ermitteln und diese von den Marketing-Mix-Effekten zu trennen. Der Ansatz wird anhand zweier empirischer Studien illustriert. Für die Produktkategorien Margarine (in den USA) und Waschmittel (in Deutschland) kommen jeweils Haushaltspaneldaten mit Beobachtungszeiträumen von 117 bzw. 226 Wochen zum Einsatz. Die Ergebnisse belegen Markenwertdynamiken für beide Produktkategorien. Aus den Verläufen lassen sich wertvolle Managementimplikationen ableiten, die bei einer statischen Modellierung im Verborgenen bleiben würden.

Es werden mehrere *Ziele* verfolgt:

1. Der eingesetzte *zweistufige FE-Ansatz* ist vielversprechend, wird aber in der Literatur (noch) nicht sehr häufig eingesetzt. Folglich soll das Verständnis hierfür verbessert werden, insb. mit Blick auf Anwendungen mit zeitvariablen Parametern.
2. Zudem gilt es, *Markenwerte* und *deren Dynamiken* generell besser zu verstehen. Hier stellen sich nämlich die Fragen, ob Markenwerte überhaupt variieren und wenn ja, wie stark.
3. Das Potenzial von *Haushaltspaneldaten* in Kombination mit dem vorgestellten Modellierungs- und Schätzansatz ist für eine *dynamische*

⁷ Im vorliegenden Beitrag wird R (R Core Team 2013) mit den Paketen *mlogit* (Croissant 2013) und *d1m* (Petrakis *et al.* 2009) genutzt. Alternativ ist aber auch etablierte Ökonometrie- bzw. Statistiksoftware, wie z. B. Stata oder SAS, einsetzbar.

Markenwertmessung bisher nicht hinreichend bekannt. Daher soll insb. Forschern in der Praxis, die mit der Markenwertmessung betraut sind, ein Einblick gegeben werden, der ihnen den Zugang hierzu erleichtert.

Der vorliegende Beitrag besitzt einen *Aufbau* mit fünf Abschnitten: In Abschnitt 4.2 wird zunächst die für das Thema relevante Literatur strukturiert und diskutiert. Hiernach erklärt Abschnitt 4.3 das eingesetzte *Modell*; die dazugehörige *zweistufige* Schätzung wird in Abschnitt 4.4 beschrieben. Abschnitt 4.5 präsentiert *zwei empirische Studien*, die detailliert illustrieren, welche Vorzüge eine dynamische nutzenbasierte Markenwertmessung besitzt. Der Beitrag endet in Abschnitt 4.6 mit einer *Zusammenfassung* der Kernergebnisse und einem *Ausblick* auf offene Forschungsfragen.

4.2 Literaturüberblick

Der *Literaturüberblick* in diesem Abschnitt betrachtet Arbeiten, die 1.) Dynamik in diskreten Wahlmodellen behandeln und/oder 2.) den FE-Ansatz von Berry (1994) und Berry *et al.* (2004) verwenden. Es werden Grenzen der aktuellen Forschung aufgedeckt und der vorliegende Beitrag entsprechend eingeordnet. Der inhaltliche Schwerpunkt liegt auf einer nutzenbasierten Messung des Markenwerts, und aus den einleitend genannten Gründen kommen nur Beiträge in Betracht, die disaggregierte Daten einsetzen.⁸

Tabelle 4.2 zeigt, dass, mit Ausnahme von Baumgartner (2003), generell eine *heterogene* Modellierung stattfindet. Mela *et al.* (1997) und Heilman *et al.* (2000) verwenden hierzu einen Ansatz mit diskreter Heterogenität, der Konsumenten während der Schätzung segmentiert (Kamakura und Russell 1989). Einige Quellen beruhen auf dem *Logit*modell, obwohl manche Autoren auch *Probit*modelle nutzen. Die Normalverteilungsannahme des Probitmodells erschwert zwar die Berechnung von Wahlwahrscheinlichkeiten, die Berücksichtigung normalverteilter Heterogenität ist hier aber einfacher.

⁸ Für Beiträge, die auf aggregierten Daten fußen, sei auf Sriram *et al.* (2007) sowie Hernández-Mireles und Fok (2011) verwiesen.

Tabelle 4.2: Ausgewählte Beiträge zur Abbildung von Dynamik in diskreten Wahlmodellen für Individualdaten

Autor(en) (Jahr)	Wahl-Modell	unbeob. Heterogenität	Dynamische Größe*	Modellierungs-/Schätzansatz für Dynamik	betrachteter Zeitraum
Papatla und Krishnamurthi (1996)	Probit	ja	Ma.-Mix-Parameter	Reparametrisierung	104 Wochen
Mela <i>et al.</i> (1997)	Logit	ja (2 Segmente)	Marg. Preis- und Promotioneffekte	1. Stufe: Moving Window 2. Stufe: AR(1)-Regression	8 1/4 Jahre
Jedidi <i>et al.</i> (1999)	Probit	ja	Nutzenparameter	Reparametrisierung	8 1/4 Jahre
Heilman <i>et al.</i> (2000)	Logit	ja (3 Segmente)	Nutzenparameter	Reparametrisierung	208 Wochen
Baumgartner (2003)	Logit	nein	Markenparameter	Cubic Splines	52 Wochen
Lachaab <i>et al.</i> (2006)	Probit	ja	Nutzenparameter	Kalman Filter	91 Monate
Rutz und Sonnier (2011)	Probit	ja	Markenfaktor	Kalman Filter	4 Jahre
Gordon <i>et al.</i> (2013)	Logit	ja	Preisparameter und -elastizitäten	1. Stufe: Quartalsindikatoren 2. Stufe: Regression	6 Jahre
Der vorliegende Beitrag	Logit	ja	Markenparameter	1. Stufe: Fixe Effekt 2. Stufe: Kalman Filter	117 und 226 Wochen

Anmerkung: (*) Als Nutzenparameter werden sämtliche Parameter in der Nutzenfunktion bezeichnet. Mit Marketing-Mix-, Preis- und Markenparameter sind die entsprechenden Untergruppen aller Nutzenparameter gemeint.

Die Arbeiten spezifizieren verschiedene Modellgrößen dynamisch. Teilweise sind sämtliche Nutzenparameter zeitvariabel (z. B. Jedidi *et al.* 1999), je nach Untersuchungsgegenstand werden manchmal aber auch nur Marketing-Mix- (Papatla und Krishnamurthi 1996) oder nur Markenparameter dynamisch modelliert (Baumgartner 2003). Von besonderem Interesse ist, wie die *Dynamik* modelliert wird. Zunächst lassen sich ein- (z. B. Lachaab *et al.* 2006) und zweistufige Ansätze (z. B. Mela *et al.* 1997) trennen. In *einstufigen* Ansätzen werden i. d. R. dynamischen Modellgrößen als Funktion zeitvariabler Kovariaten („process functions“, Leeﬂang *et al.* 2000) *reparametrisiert*. Heilman *et al.* (2000) erklären zeitliche Parameterunterschiede z. B. durch die Kategorieerfahrung der Konsumenten, welche wiederum als Anzahl der Kategoriekäufe operationalisiert wird. Das Vorgehen ist sehr einfach und mit Standardsoftware durch Interaktionsterme umsetzbar. Modelliert wird hierbei aber nur die zeitliche Varianz der dynamischen Variablen und nicht deren Verläufe, weil lediglich die Erwartungswerte bedingt auf die erklärenden Variablen gebildet werden können (van Heerde *et al.* 2004). Zeitlich Parameterverläufe erhält man dagegen durch *nichtparametrische Regressionen* („Splines“, Baumgartner 2003 und allgemein z. B. Fahrmeir *et al.* 2013) oder mithilfe des KF (Lachaab *et al.* 2006 bzw. Rutz und Sonnier 2011 und allgemein Harvey 1989 sowie Kapitel 2 der vorliegenden Arbeit). Diese Verfahren bieten eine größere Flexibilität und insb. der KF gehört heutzutage für die Modellierung dynamischer Zusammenhänge zu den State of the Art-Methoden im Marketing (Leeﬂang *et al.* 2009). In Kombination mit einstufigen Verfahren führen diese Methoden aber zu einem sehr hohen Schätzaufwand.

Zweistufige Verfahren schätzen auf der ersten Stufe das diskrete Wahlmodell und ermitteln hierbei Größen, die auf der zweiten Schätzstufe abhängige Variablen darstellen. Mela *et al.* (1997) schätzen z. B. einzelne marginale Preis- und Promotioneffekte für alle Quartale (33 Stück), indem sie eine *Moving Window-Analyse* über einen langen Zeitraum mit Daten aus jeweils drei Quartalen durchführen. Pro „Fenster“ wird also ein Wahlmodell mit einem spezifischen Datenausschnitt geschätzt. Auf der zweiten Stufe folgt eine Regression mit einem Carry Over-Term zur Abbildung der Dynamik auf Ebene der marginalen Effekte als abhängige Variable. Gordon *et al.* (2013) verwenden zunächst *Quartalsindikatoren* (24 Stück), um zeitvariable Preisparameter zu erhalten. Danach regressieren sie die in Elastizitäten umgerechneten Preisparameter aus mehreren Kategorien auf makroökonomische Variablen (z. B. das Wachstum des Bruttoinlandsprodukts). Das Vorgehen bei Gordon *et al.* (2013) wird also durch die gepoolte Analyse über Kategorien hinweg begründet, ansonsten wäre direkt eine Reparamet-

trisierung möglich. Der vorliegende Beitrag schätzt zunächst *Marken-FE*. Anschließend kommt der KF zum Einsatz, um zeitvariable Markenwerte flexibel und sparsam parametrisiert aus den FE zu ermitteln.

FE-Ansätze werden immer populärer und sind konzeptionell leicht verständlich (Berry *et al.* 2004): Zunächst werden FE geschätzt, die Variationen in Dimensionen aufnehmen, in denen spezifische empirische Phänomene Schwierigkeiten bei der Schätzung des Wahlmodells bereiten (z. B. Preisendogenität, vgl. Chintagunta *et al.* 2005). Diese FE lassen sich dann als abhängige Variablen mit Standardverfahren „außerhalb“ des Wahlmodells leichter handhaben. Im vorliegenden Beitrag wird der FE-Ansatz erstmalig als Grundlage zur Schätzung dynamischer Markennutzen eingesetzt.

Tabelle 4.3 fasst bisherige FE-Beiträge für Individualdaten zusammen. Bis auf Walker *et al.* (2011) verwenden alle Arbeiten *heterogene Modelle* mit einer kontinuierlichen Heterogenitätsspezifikation. Die „frühen“ Artikel, wie z. B. Chintagunta *et al.* (2002, 2005) sowie Chintagunta und Dubé (2005), sind aber aufgrund der extrem geringen Anzahl der Zufallszüge zur Berechnung der Heterogenitätsintegrale zu kritisieren. Bis auf Goolsbee und Petrin (2004) kommt immer das *Logitmodell* zum Einsatz, was sich durch die leichte Berechnung der Wahlwahrscheinlichkeiten erklären lässt. Arbeiten, die Marken- oder Produktwahl behandeln, verwenden typische *Haushaltspaneldaten* mit einer dreistelligen Anzahl von Haushalten (429 bis 992), einer geringen Anzahl von Marken bzw. Alternativen (2 bis 4, Chintagunta und Dubé (2005) bilden mit 25 Alternativen ein Ausnahme) und einer hohen Anzahl von Wochen (82 bis 117).

Die Datenstruktur der Artikel ohne direkten Marketingbezug (Goolsbee und Petrin 2004; Berry *et al.* 2004; Murdock 2006; Train und Winston 2007; Petrin und Train 2010 und Walker *et al.* 2011) weisen mehrere Unterschiede auf. Auffällig sind einerseits die *Querschnittsdaten*, die sich u. U. über mehrere geografische Märkte erstrecken. Andererseits ist die Anzahl der untersuchten Haushalte besonders hoch (teilweise im fünfstelligen Bereich). Zur Berechnung der FE wird entweder der *Contraction Mapping*-Algorithmus von Berry *et al.* (1995)⁹ eingesetzt oder es erfolgt direkt eine Maximum-Likelihood-Schätzung (Chintagunta *et al.* 2005). Letzteres ist aufwendig, weil der Parameterraum um die Anzahl der FE steigt, welche i. d. R. im dreistelligen Bereich liegt. Der Contraction Mapping-Algorithmus erleichtert die Berechnung der FE, es gibt jedoch hierfür keine Standardsoftware (Murdock 2006; Petrin und Train 2010).

⁹ Vgl. hierzu auch Abschnitt 3.3.1 der vorliegenden Arbeit.

Tabelle 4.3: Ausgewählte Beiträge zur Verwendung fixer Effekte in diskreten Wahlmodellen für Individualdaten

Autor(en) (Jahr)	Auswahl- gegenstand	Wahl- Modell	Anzahl der HH	Anzahl der FE ($= J \cdot T$) ^{***}	Berechnung der FE	unbeob. Het. (# Züge)
Chintagunta <i>et al.</i> (2002) [*]	Orangensaftmarken	Logit	429	$2 \cdot 117 = 234$	ML	ja (30)
Berry <i>et al.</i> (2004)	Autos	Logit	37500	$200 \cdot 1 = 200$	CM (SA)	ja (NA) ^{**}
Goolsbee und Petrin (2004)	TV-Empfangsarten	Probit	29484	$3 \cdot 317 = 951$ ^{****}	NLS (SA)	ja (4000)
Chintagunta <i>et al.</i> (2005)	Margarinemarken	Logit	992	$4 \cdot 105 = 420$	ML	ja (30)
Chintagunta und Dubé (2005)	Weichspülerprodukte	Logit	657	$25 \cdot 82 = 2050$	CM (MA)	ja (30)
Murdock (2006)	Angelplätze	Logit	512	$568 \cdot 1 = 568$	CM (SA)	ja (NA) ^{**}
Train und Winston (2007)	Autos	Logit	458	$200 \cdot 1 = 200$	CM (MA)	ja (200H)
Park und Gupta (2009)	Papiertüchermarken	Logit	880	$4 \cdot 103 = 412$	ML	ja (NA) ^{**}
Petrin und Train (2010)	TV-Empfangsarten	Logit	11810	$3 \cdot 172 = 516$ ^{****}	CM (SA)	ja (100H)
Walker <i>et al.</i> (2011)	Verkehrsmittel	Logit	2913	$2 \cdot 77 = 144$ ^{****}	ML	nein
Der vorliegende Beitrag	Margarinemarken	Logit	992	$4 \cdot 105 = 420$	ML	ja (500H)
Der vorliegende Beitrag	Waschmittelmarken	Logit	2055	$4 \cdot 165 = 660$	ML	ja (500H)

Anmerkung: (*) Chintagunta *et al.* (2002) ist die Arbeitspapierversion von Chintagunta *et al.* (2005), die zusätzlich die Orangensaftkategorie enthält. (**) Die Anzahl der Züge wird nicht genannt. (***) J ist die Anzahl der Wahlalternativen im Modell und T bezeichnet die Anzahl der Perioden (i. d. R. Wochen). (****) T bezieht sich hier nicht auf Perioden sondern auf geografische Märkte. CM (MA) = Contraction Mapping mit Marktanteilen; CM (SA) = Contraction Mapping mit Sampleanteilen; H = Halton; HH = Haushalte; ML = Maximum-Likelihood; NLS (SA) = Nonlinear Least Squares mit Sampleanteilen.

Der vorliegende Beitrag passt hinsichtlich der in Tabelle 4.3 aufgeführten Studienmerkmale gut zu den restlichen Arbeiten. Zwei Unterschiede sind jedoch auffällig: 1.) Es kommen 500 *Haltonzüge* zur Integration über die unbeobachtete Heterogenität zum Einsatz. Das entspricht dem heutigen Stand in der Forschung und sorgt für einen vernachlässigbaren Simulationsbias (Train 2009). 2.) Weil es sich beim Untersuchungsgegenstand um zeitvariable Markenwerte handelt, ist ein besonders *großer Untersuchungszeitraum* vorteilhaft. Pro Marke gibt es in der Waschmittelkategorie 165 *wöchentliche FE*, was überdurchschnittlich viel ist.¹⁰

Zusammenfassend kann festgehalten werden, dass sich der vorliegende Beitrag am aktuellen Stand der Forschung ausrichtet. Durch eine innovative Anwendung des FE-Ansatzes wird eine einfache Schätzung dynamischer Markenwerte mit dem KF ermöglicht. *Erweiterungen* des vorliegenden Beitrags können im Sinne einer *evolutionären Modellbildung* verstanden werden, wie sie van Heerde *et al.* (2002) anregen. Bei dieser Vorgehensweise wird zunächst relativ einfach begonnen, um Probleme in der Modellierung zu identifizieren und zu eliminieren. Hiernach erfolgt dann Schritt für Schritt eine Steigerung der Komplexität.

Seit der wegweisenden Arbeit von Guadagni und Little (1983) sind diskrete Markenwahlmodelle im Marketing sehr verbreitet. In den 1990er Jahren ging es vorrangig um Konsumentenheterogenität (vgl. z. B. Jain *et al.* 1994), Ende der 1990er Jahre und in den 2000er Jahren wurde Preisendogenität thematisiert (vgl. z. B. Villas-Boas und Winer 1999) und inzwischen besteht ein großes Interesse an zeitvariablen Effekten (vgl. z. B. Rutz und Sonnier 2011). Betrachtet man nur diskrete Wahlmodelle mit Dynamik, so ist dort ebenfalls ein evolutionärer Prozess zu identifizieren, beginnend mit Re-parametrisierungen, über mehrstufige Ansätze, bis hin zu Zustandsraummodellen und dem KF. Diese generelle Modellevolution wird unterstützt durch immer reichhaltigere Datenquellen, größerer Rechenleistung moderner Computer und verbesserter Software (Leeftang *et al.* 2009). Zudem erlauben verbesserte Schätzalgorithmen und -strategien die Anwendung komplexerer Modelle (vgl. z. B. Berry 1994).

Vor diesem Hintergrund erscheint das hier gewählte innovative Vorgehen, einer Schätzung von FE auf der ersten Stufe und einer anschließenden Anwendung des KF zur Bestimmung dynamischer Markennutzen auf der zweiten Schätzstufe, als logische Konsequenz. Einzelheiten zum Aufbau

¹⁰ Der gesamte Zeitraum beträgt 226 Wochen, wegen fehlender Werte können aber nicht für alle Wochen FE geschätzt werden.

und zur zweistufigen Schätzung des verwendeten Modells sind in den beiden nächsten Abschnitten aufgeführt.

4.3 Modellierung

4.3.1 Nachfragemodell

Grundlage für die *Nachfragemodellierung* ist ein diskretes Markenwahlmodell, wie es in der Marketingforschung seit Jahren häufig eingesetzt wird (Guadagni und Little 1983; Jain *et al.* 1994; Villas-Boas und Winer 1999; Dubé *et al.* 2010; Gordon *et al.* 2013). Konsumenten führen Einkäufe bei einem Händler durch und entscheiden sich in einer spezifischen Kategorie entweder für eine der verfügbaren Marken oder gegen einen Kauf.¹¹ Letzteres sei als „Nichtwahloption“ bezeichnet und ermöglicht die Modellierung der Kategorienachfrage.

Für die (indirekte) Nutzenfunktion gilt:

$$u_{ijt} = \alpha_{ijt} + \mathbf{x}'_{jt} \boldsymbol{\beta}_i + \lambda_i \text{sd}_{ijt} + \varepsilon_{jt} + \nu_{ijt}, \quad \text{mit} \quad (4.1) \\ i = 1, \dots, I, \quad j = 1, \dots, J \text{ und } t = 1, \dots, T.$$

Der Gesamtnutzen u_{ijt} von Person i aus der Wahl einer Marke j zum Zeitpunkt t setzt sich aus mehreren Einflussgrößen zusammen, während für die Nichtwahloption ($j = 0$) vereinfachend $u_{i0t} = \nu_{i0t}$ gilt.

α_{ijt} ist der *individuelle* und *zeitvariierende* Nutzen der Marke j , der nicht auf Marketing-Mix- oder Kauffeedbackeffekte zurückführbar ist. Der Spaltenvektor \mathbf{x}_{jt} beinhaltet marken- und zeitspezifische *Marketing-Mix*-Variablen, die für alle Personen gleich sind. Hierzu gehört auch der in seiner Wirkung linear spezifizierte *Preis* p_{jt} .¹² $\mathbf{x}'_{jt} \boldsymbol{\beta}_i$ ist somit der aus dem Marketing-Mix resultierende Nutzen, welcher ebenfalls personenspezifisch ist, da die Marketing-Mix-Sensitivitäten $\boldsymbol{\beta}_i$ als heterogen angesehen werden.

¹¹ Der Einkaufszeitpunkt und -ort sind modellexogen und die Mengenwahl wird nicht betrachtet. Diese Punkte stellen interessante Erweiterungen für zukünftige Arbeiten dar.

¹² Durch diese Spezifikation des Preises und das Vorhandensein der Nichtwahloption lassen sich Nutzenwerte leicht in monetäre Einheiten umrechnen („Zahlungsbereitschaften“), weil dann der Preisparameter in der Nutzenfunktion als *marginaler Nutzen des Einkommens* interpretierbar ist (Dubé *et al.* 2010).

Guadagni und Little (1983), Keane (1997) und Dubé *et al.* (2010) belegen empirisch, dass Konsumenten mit einer höheren Wahlwahrscheinlichkeit Marken wählen, die sie bereits in der Vergangenheit gekauft haben. Hierdurch entsteht ein typisches Einkaufsmuster, bei welchem zwar Markenwechsel vorkommen, nach einem Wechsel aber bei derselben Marke verharret wird. Diese *Verharrungstendenz* der Konsumenten („inertia“, Keller und Lehmann 2006, S. 751) wird in der Literatur als *Strukturelle State Dependence* bezeichnet, wenn sie direkt durch *vergangene Käufe* ausgelöst wird (Heckman 1981), und lässt sich z. B. dadurch begründen, dass Konsumenten ihre Suchkosten minimieren. Der durch vergangene Käufe induzierte Nutzen wird mit dem personenspezifischen Parameter λ_i abgebildet. Die Konsumenten im Modell dürfen also unterschiedlich stark an ihren vergangenen Markenwahlentscheidungen festhalten.¹³ Die Kaufhistorie der Konsumenten wird über die State Dependence-Variable sd_{ijt} operationalisiert. In Anlehnung an Dubé *et al.* (2010) gilt $sd_{ijt} = \mathbb{I}(\text{Letztkauf}_{it} = j)$. Somit ist sd_{ijt} eine Dummyvariable, die nur für jene Marke j zum Zeitpunkt t den Wert Eins annimmt, wenn diese Marke j beim letzten Kategoriekauf gewählt wurde (und ansonsten den Wert Null annimmt).¹⁴

Der State Dependence-Effekt wird in der Marketingliteratur oftmals als *Loyalität* bezeichnet (Guadagni und Little 1983). Vor dem Hintergrund des im vorliegenden Beitrag behandelten Themas, ist dieser Terminus jedoch missverständlich und wird aus diesem Grund nicht verwendet. Die Loyalität für eine (bestimmte) Marke, ist nämlich ein *Markenwertindikator* bzw. „*Brand Value Driver*“ auf Konsumentenebene (Sattler 2005; Keller und Lehmann 2006), und ist folglich ist der Effekt der Markenloyalität im Markenwert enthalten. Ein höherer Markenwert für Marke j führt *ceteris paribus* immer zu einer durchweg höheren Wahlwahrscheinlichkeit für eben diese Marke. Aus der Verharrungstendenz dagegen folgt nur dann eine erhöhte Wahlwahrscheinlichkeit, wenn zuvor auch Marke j gewählt wurde und hierfür mag es wie erwähnt verschiedene Gründe geben. Die Ausführungen verdeutlichen, dass die explizite Berücksichtigung von Kauffeedbackeffekten für die präzise und verzerrungsfreie Markenwertmessung sehr

¹³ Es ist sogar denkbar, dass für einige Konsumenten $\lambda_i < 0$ gilt. Jenes kann z. B. der Fall sein, wenn diese Konsumenten ein besonders ausgeprägtes *Abwechslungsbedürfnis* aufweisen (Chintagunta 1998; Seetharaman und Chintagunta 1998).

¹⁴ Alternativ könnte sd_{ijt} auch als exponentiell geglättete Variable spezifiziert werden (Guadagni und Little 1983). Hierdurch würden *alle* zurückliegenden Käufe einen Effekt auf den aktuellen Nutzen besitzen. Im vorliegenden Beitrag wird hiervon aber abgesehen, weil sonst einerseits auch noch zusätzlich der Glättungsparameter zu bestimmen wäre und andererseits mehr Daten zur Variableninitialisierung „geopfert“ werden müssten.

wichtig ist (Keller und Lehmann 2006, S. 751). Nur Haushaltspaneldaten mit langem Zeithorizont (z. B. mehr als 52 Wochen) sind für eine Trennung beider Effekte ideal geeignet (Dubé *et al.* 2009).

Für den Analytiker *unbeobachtete Effekte* werden durch die zwei Störterme ε_{jt} und ν_{ijt} modelliert. ν_{ijt} ist *unabhängig identisch Typ I extremwertverteilt* und bildet Einflüsse ab, die über Marken, Zeitpunkte und vor allem Personen variieren. Zum Beispiel könnte eine Person beim Händler aus Unaufmerksamkeit eine eigentlich nicht beabsichtigte Marke kaufen. ε_{jt} ist *normalverteilt* mit einem Mittelwert von Null und einer Varianz von σ_ε^2 und bildet unbeobachtete Umstände ab, die den Nutzen für alle Personen gleichermaßen beeinflussen, also nur über Marken und Zeitpunkte variieren. Als Beispiel hierzu lässt sich die Sonderplatzierung des Händlers oder Herstellerwerbung anführen, diese werden nachfolgend, wie in der Literatur üblich, als UMC bezeichnet (Chintagunta *et al.* 2005).

Für die *heterogenen Parameter* gilt die Annahme, dass sie in der Population (= Käuferschaft) *multivariat normalverteilt* sind. Diese Verteilungsannahme ist in der Marketingliteratur weit verbreitet und ermöglicht eine flexible Modellierung komplexer Substitutionsmuster (Chintagunta *et al.* 2005; Erdem *et al.* 2008):

$$\begin{bmatrix} \alpha_{it} \\ \beta_i \\ \lambda_i \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_t \\ \bar{\beta} \\ \bar{\lambda} \end{bmatrix} + \begin{bmatrix} \nu_{\alpha i} \\ \nu_{\beta i} \\ \nu_{\lambda i} \end{bmatrix}, \text{ mit } \begin{bmatrix} \nu_{\alpha i} \\ \nu_{\beta i} \\ \nu_{\lambda i} \end{bmatrix} = \mathbf{v}_i \sim N(\mathbf{0}, \Psi). \quad (4.2)$$

Gemäß (4.2) variieren die J Mittelwerte der Markennutzen (die „Durchschnittsmarkennutzen“) über die Zeit. Die Heterogenität als Abweichung vom Durchschnitt bleibt dagegen stabil (Lachaab *et al.* 2006). Hierdurch wird der Überlegung Rechnung getragen, dass „echte“ Änderungen im Markenwert gerade nicht auf individuelle Effekte wie State Dependence ($\lambda_i \text{sd}_{ijt}$) zurückzuführen sind und keine situativen Ursachen besitzen (ν_{ijt}). Weil im Modell gleichzeitig Marketing-Mix- und UMC-Effekte berücksichtigt werden, ist $\bar{\alpha}_{jt}$ der durchschnittliche Nutzen einer Marke, der keine andere Ursache als die Marke selbst hat. Dieser von Keller und Lehmann (2006) als „*incremental value*“ bezeichnete Term entspricht dem *Markenwertmaß*¹⁵ von Kamakura und Russell (1993) mit der im Rahmen des vor-

¹⁵ Es sei betont, dass in Anlehnung an Kamakura und Russell (1993) im vorliegenden Beitrag auch dann der Terminus *Markenwert* verwendet wird, wenn dieser *nicht* monetär skaliert ist, sondern abstrakt in Nutzeneinheiten. Wie zuvor in Fußnote 12 erwähnt, ist ein Umrechnung des Nutzens in Geldeinheiten aber leicht möglich.

liegenden Beitrags wichtigen ergänzenden Annahme der *Zeitvariabilität*.

Man beachte, dass *absolute Nutzenhöhen* in diskreten Wahlmodellen *nicht interpretierbar* sind, sondern nur Nutzendifferenzen Aussagekraft besitzen (Train 2009). Das Modell ist *identifiziert*, da der Nutzen für die Nichtwahl-option (willkürlich) auf Null gesetzt ist. Für eine gegenüberstellende *Markenwertinterpretation* eignet sich die Nichtwahloption jedoch nicht, da es sich hierbei nicht um ein real existierendes Produkt handelt. In der Regel wird der Vergleich zu einer markenlosen Alternative oder einer Handelsmarke gezogen (Sattler 2005). Solche fiktiven Vergleiche sind im postulierten Modellrahmen sehr einfach durchführbar und passen sogar zur notwendigen Differenzenbetrachtung. Angenommen $j = 1$ sei die Handelsmarke, dann entspricht $\hat{\alpha}_{jt} := \bar{\alpha}_{jt} - \bar{\alpha}_{1t}$ dem Markenwert von j im Vergleich zur Handelsmarke (Ailawadi *et al.* 2003). Wegen der dynamischen Markenwertbetrachtung ist auch ein zeitlicher Vergleich möglich und es könnte ein bestimmter Zeitpunkt zur Verankerung dienen. Zum Beispiel bezeichnet $\check{\alpha}_{jt} := \bar{\alpha}_{jt} - \bar{\alpha}_{j1}$ den Markenwert von j zum Zeitpunkt t im Vergleich zu $t = 1$ (Hernández-Mireles und Fok 2011).

Die Dynamik in den j Markenwerten wird für $t = 1, \dots, T$ über *Random Walk-Prozesse* modelliert (Lachaab *et al.* 2006; Rutz und Sonnier 2011):

$$\bar{\alpha}_t = \bar{\alpha}_{t-1} + \boldsymbol{\eta}_t, \text{ mit } \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}). \quad (4.3)$$

Solche stochastischen und nichtstationären Dynamiken sind vorteilhaft, weil sich hierdurch Markenwerte im Modell völlig frei entwickeln dürfen und keine Vorgabe für spezifische Entwicklungen erfolgt (Putsis 1998). Die Markenwertverläufe sind nicht durch eine funktionale Form oder erklärende Größen prädestiniert, sondern ergeben sich ausschließlich aus den Daten („parameter-driven dynamics“, vgl. Lachaab *et al.* 2006). Dadurch wird dem Anspruch einer adäquaten dynamischen Messung eines empirischen Phänomens nachgekommen, für welches a priori kein oder nur sehr wenig Vorwissen existiert. Zudem reicht es aus, die Parameter der Kovarianzmatrix \mathbf{Q} zu schätzen, welche misst, wie stark Schwankungen im Markenwert von Zeitpunkt zu Zeitpunkt ausfallen.¹⁶ Aus Gründen einer sparsamen Parametrisierung wird $\mathbf{Q} = \sigma_{\eta}^2 \cdot \mathbf{I}_j$ gewählt, wobei \mathbf{I}_j eine j -dimensionale Einheitsmatrix darstellt. Für $\sigma_{\eta}^2 > 0$ gibt es Dynamik in den Markenwerten, $\sigma_{\eta}^2 = 0$ führt dagegen zu intertemporal konstan-

¹⁶ Schätzungen für $\boldsymbol{\alpha}_t$ lassen sich über den KF ermitteln und nehmen keine Freiheitsgrade in Anspruch.

ten Markenwerten ($\bar{\alpha}_t = \bar{\alpha}$). Deshalb ist ein Test auf Dynamik direkt über diesen Parameter möglich.

Die getroffenen Verteilungsannahmen für die Heterogenität erlauben eine kompaktere Darstellung der Nutzenfunktion (4.1):

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + v_{ijt}, \text{ mit} \quad (4.4)$$

$$\delta_{jt} = \bar{\alpha}_{jt} + \mathbf{x}'_{jt} \bar{\boldsymbol{\beta}} + \varepsilon_{jt} \text{ und} \quad (4.5)$$

$$\mu_{ijt} = \bar{\lambda} \text{sd}_{ijt} + \left[\mathbf{1}_j \mathbf{x}'_{jt} \text{sd}_{ijt} \right] \boldsymbol{\Psi}^{1/2} \mathbf{v}_i^* \text{ für} \quad (4.6)$$

$$i = 1, \dots, I, j = 1, \dots, J \text{ und } t = 1, \dots, T.$$

Diese Schreibweise verdeutlicht, welche Modellbestandteile personenspezifisch sind (μ_{ijt}) und welche nicht (δ_{jt}). δ_{jt} stellt den *Durchschnittsnutzen* einer Marke j zum Zeitpunkt t dar und beinhaltet alle Populationseffekte inkl. zeitvariabler Markenwerte und UMC. Genau diese $\boldsymbol{\delta} = \{\delta_{jt}\}_{j=1,t=1}^{J,T}$ sind die zuvor erwähnten FE. μ_{ijt} lässt sich als heterogene *Abweichung vom Durchschnittsnutzen* interpretieren, wobei $\boldsymbol{\Psi}^{1/2}$ die Cholesky-Wurzel (untere Dreiecksmatrix) von $\boldsymbol{\Psi}$ darstellt und \mathbf{v}_i^* dementsprechend standardnormalverteilt ist (Chintagunta *et al.* 2005). In μ_{ijt} wird v_{ijt} nicht mit aufgenommen, da sich hierüber bequem geschlossen integrieren lässt, um individuelle *Wahlwahrscheinlichkeiten* zu erhalten. Wenn die Personen zu allen Zeitpunkten immer genau die Marke wählen, die ihnen den höchsten Nutzen stiftet, folgen die Wahlwahrscheinlichkeiten dem bekannten *mixed Logitmodell* (MXL-Modell) (Train 2009):

$$\text{pr}_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{j' \in J} \exp(\delta_{j't} + \mu_{ij't})}, \text{ mit} \quad (4.7)$$

$$i = 1, \dots, I, j = 1, \dots, J \text{ und } t = 1, \dots, T.$$

Die Eins im Nenner von (4.7) ergibt sich, weil die Nichtwahloption einen Nutzen von Null hat. Die Formel deckt auf, dass im Logitmodell die Wahlwahrscheinlichkeiten der Marken und der Nichtwahloption von den Gesamtnutzenwerten sämtlicher Wahlalternativen abhängen. Somit gibt es im Modell implizite Konkurrenzeffekte und diese konfundieren nicht die Schätzung der Markenwerte. Würde Konkurrenz vernachlässigt, könnte ein fallender Absatztrend einer Marke – isoliert betrachtet – auf einen Verlust des Markenwerts hinweisen. Die tatsächliche Ursache mag aber sein, dass ein Konkurrent bei konstanter Marktgröße mittels seiner Aktivitäten

stetig Marktanteile gewinnt. Die diskutierte Modelleigenschaft hat deshalb eine wichtige Implikation für die Nichtwahloption. Auch wenn keine Faktoren Berücksichtigung finden, die direkt auf den Nutzen der Nichtwahloption wirken (z. B. Lagerhaltungseffekte), steuern die Nutzen aller Marken auch die Wahlwahrscheinlichkeit für die Kategoriewahl. Günstige Preise, Verkaufsförderungsaktionen oder eben steigende Markenwerte führen z. B. dazu, dass Konsumenten *ceteris paribus* eher auf die Nichtwahl verzichten.

4.3.2 Endogenitätskorrektur

Preisendogenität ist in der Markenwahlmodellierung mit nichtexperimentellen Marktdaten ein bedeutsames Thema (Villas-Boas und Winer 1999; Kuksov und Villas-Boas 2008). Preisendogenität liegt vor, wenn Preise und UMC korreliert sind ($E[\varepsilon_{jt}|p_{jt}] \neq 0$). Dies ist eine naheliegende Vermutung, wenn man sich die inhaltliche Bedeutung der UMC verdeutlicht. Es erscheint unplausibel, dass Markencharakteristika, die nur für den Analytiker unbeobachtbar sind, vom Anbieter in der Preissetzung unberücksichtigt bleiben. In aller Regel sollten nutzenstiftende Markeneigenschaften (z. B. gute Regalplatzierung, nationale Werbekampagnen, o. ä.) mit höheren Preisen einhergehen. Dies impliziert eine positive Korrelation zwischen Preisen und UMC, wodurch (absolut gesehen) zu kleine Preissensitivitäten folgen. Diese Richtung der Endogenitätsverzerrung belegen z. B. Bijmolt *et al.* (2005) in ihrer empirischen Verallgemeinerung zu Preiselastizitäten. Das Ignorieren der UMC in Wahlmodellen – bis heute in der Literatur üblich – trägt nicht dazu bei, das Problem zu lösen, weil dann andere Modellkomponenten die Effekte der UMC aufnehmen (Park und Gupta 2009).

Für die nutzenbasierte Messung der Markenwerte ist die Berücksichtigung der Preisendogenität aus zwei Gründen von zentraler Bedeutung:

1. Wird die Preissensitivität der Konsumenten falsch ermittelt, sind i. d. R. auch die Markennutzen verzerrt.
2. Selbst wenn letzteres nicht der Fall sein sollte, sind aufgrund der falsch ermittelten Preissensitivität alle hiervon abgeleiteten Größen verzerrt. Dazu gehört u. a. das monetäre Äquivalent des Markenwerts, nämlich die Zahlungsbereitschaft für Marken.

Im vorliegenden Beitrag kommen zur Behandlung des Endogenitätspro-

blems sog. *Kontrollfunktionen* (CF) zum Einsatz (Park und Gupta 2009; Petrin und Train 2010). Der dazugehörige CF-Ansatz arbeitet wie folgt:¹⁷ Der möglicherweise endogene Preis wird auf *Instrumentalvariablen* (\mathbf{z}_{jt}) regressiert und der Störterm dieser Regression (ζ_{jt}) fängt Preisvariationen auf, die nicht exogen zu erklären sind (Pancras 2010; Gordon *et al.* 2013):

$$\begin{aligned} p_{jt} &= \mathbf{z}_{jt}' \boldsymbol{\gamma} + \zeta_{jt}, \text{ mit} \\ j &= 1, \dots, J \text{ und } t = 1, \dots, T. \end{aligned} \quad (4.8)$$

Sind UMC und ζ_{jt} korreliert, liegt ein Endogenitätsproblem vor. Die Ergänzung von $\zeta_{jt}^* = \zeta_{jt} / \sigma_\zeta$ in der Nutzenfunktion als zusätzliche Variable fängt genau den Teil der Nutzenvariation auf, der ansonsten irrtümlicherweise der Preissensitivität zugeordnet wird. Weil die Endogenität nur die FE betrifft, ist also lediglich (4.5) zu modifizieren (Berry 1994; Chintagunta *et al.* 2005):

$$\delta_{jt} = \bar{\alpha}_{jt} + \mathbf{x}_{jt}' \bar{\boldsymbol{\beta}} + \iota_j \zeta_{jt}^* + \varepsilon_{jt}. \quad (4.9)$$

Über ι_j lässt sich ohne Vorzeichenrestriktion für jede Marke ermitteln, ob Preisendogenität vorliegt (Park und Gupta 2009). Weil im Modell $\iota_j = 0$ erlaubt wird, lässt sich die statistische Signifikanz dieser CF-Parameter (also der Preisendogenität von j) direkt über t -Tests ermitteln. Durch Likelihood-Quotienten- oder Wald-Tests kann zudem die gemeinsame Nullhypothese, „Keine Marke weist ein Endogenitätsproblem auf“ ($H_0 : \{\iota_j\}_{j=1}^J = \mathbf{1} = \mathbf{0}$), überprüft werden (Park und Gupta 2009; Greene 2011).

4.4 Schätzung

Die *Maximum Likelihood* (ML)-Schätzung des Modells erfolgt in *zwei Stufen* (vgl. Abschnitt 4.2).¹⁸ Dabei bietet sich eine likelihoodbasierte Schätzung bei diskreten Wahlmodellen oftmals an, weil zur Modellherleitung ohnehin Verteilungsannahmen getroffen werden müssen (z. B. die Typ I

¹⁷ Vgl. zum Kontrollfunktionenansatz auch ausführlich Abschnitt 3.2.3 in Kapitel 3 der vorliegenden Arbeit.

¹⁸ Dieses Vorgehen unterscheidet den vorliegenden Beitrag von denen in den Kapiteln 2 und 3. Dort ist aufgrund der Verwendung *aggregierter Daten* die explizite Trennung der Schätzung in zwei Stufen *nicht möglich*.

Extremwertverteilung, die zum Logitmodell führt). Für eine zweistufige Schätzung sprechen *methodische, praktische und didaktische Gründe*:

1. Methodisch ist die Handhabung von UMC, Endogenität *und* zeitvariablen Parametern viel einfacher, da es sich auf Nutzenebene um ein lineares Modell handelt. Die Bandbreite zulässiger und etablierter Methoden ist für lineare Modelle deutlich größer.
2. Die praktische Vereinfachung liegt darin begründet, dass sich einerseits Standardsoftware für beide Stufen einsetzen lässt, z. B. R, Stata oder SAS. Das führt zu geringeren „Rüstkosten“ mit robusteren Programmen im Vergleich zu selbstprogrammierten Lösungen. Andererseits ist die getrennte Schätzung zusammen schneller als eine gemeinsame Schätzung beider Stufen.
3. Die Zweistufigkeit fördert ein tieferes Modellverständnis, denn dem Analytiker muss klar sein, welche Variation in den Daten mit ihren unterschiedlichen Dimensionen (Personen, Marken, Zeitpunkte) die spezifischen Effekte im Modell identifiziert.

Zweistufige Verfahren haben den *Nachteil* einer geringeren statistischen Effizienz. Übliche Haushaltspaneldaten weisen aber eine sehr hohe Informationsdichte auf, was auch die empirischen Studien in Abschnitt 4.5 belegen werden. In einem anwendungsorientierten Kontext ist dieser Kritikpunkt aber von *untergeordneter Bedeutung*, insb. vor dem Hintergrund mangelnder Alternativen (Chintagunta *et al.* 2005). Keiner der in Tabelle 4.2 aufgeführten Ansätze zur Modellierung von Dynamik in diskreten Wahlmodellen berücksichtigt *gleichzeitig* Heterogenität, Endogenität *und* UMC.

4.4.1 Erste Stufe

Die zuvor getroffenen Annahmen ermöglichen es, die Log-Likelihoodfunktion zur Schätzung des diskreten Wahlmodells aufzustellen. Es werden alle FE (insg. $J \cdot T$ δ_{jt} -Parameter in δ) und die Parameter in μ_{ijt} ($\bar{\alpha}$ und Heterogenitätsparameter θ_Ψ , vgl. Gleichung (4.6)) auf dieser ersten Stufe geschätzt.¹⁹ Die Schätzung ist bei einer hohen Anzahl von FE ($\gg 100$) durchaus aufwendig, mit entsprechenden Daten und heutigen Computern

¹⁹ Die Schätzung der Populationsmittelwerte der Nutzenparameter ($\bar{\alpha}_i$ und $\bar{\beta}$) findet auf der zweiten Stufe statt (vgl. Abschnitt 4.4.2).

aber problemlos zu bewältigen.

Die Wahrscheinlichkeitsdichten der Einkaufshistorien für alle Konsumenten bedingt auf die Daten und gesuchten Parameter (inkl. der FE) lauten (Chintagunta *et al.* 2005):

$$L_i(\mathbf{w}_i, \mathbf{x}, \mathbf{sd}_i, \boldsymbol{\delta}, \bar{\lambda}, \boldsymbol{\theta}_\Psi) = \int \dots \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} \prod_{j=1}^J \text{pr}_{ijt}(\mathbf{x}_{jt}, \mathbf{sd}_{ijt}, \delta_{jt}, \bar{\lambda}, \boldsymbol{\theta}_\Psi, \mathbf{v}_i^*)^{w_{ijt}} \phi(\mathbf{v}_i^* | \mathbf{0}, \mathbf{I}_J) d\mathbf{v}_i^*, \quad (4.10)$$

$$\approx \frac{1}{S} \sum_{s=1}^S \prod_{t=1}^{T_i} \prod_{j=1}^J \text{pr}_{ijt}(\mathbf{x}_{jt}, \mathbf{sd}_{ijt}, \delta_{jt}, \bar{\lambda}, \boldsymbol{\theta}_\Psi, \mathbf{v}_s^*)^{w_{ijt}}, \quad \text{mit} \quad (4.11)$$

$$w_{ijt} = \begin{cases} 1, & \text{wenn die Marke } j \text{ ausgewählt wird und} \\ 0, & \text{sonst.} \end{cases}$$

$\phi(\mathbf{v}_i^* | \mathbf{0}, \mathbf{I}_J)$ ist die Dichte einer J -dimensionalen Standardnormalverteilung und die Integration über \mathbf{v}_i^* ist notwendig, da die Heterogenität *unbeobachtet* ist. Werte hierzu sind für den Analytiker nicht zugänglich und müssen daher aus der Likelihoodfunktion „herausintegriert“ werden. Das Mehrfachintegral in (4.10) ist nicht geschlossen lösbar. Aus diesem Grund kommen Monte Carlo-Techniken zum Einsatz (Keane 1997; McFadden und Train 2000; Train 2009). In (4.11) werden $S = 500$ Haltonzüge aus $\phi(\mathbf{v}^* | \mathbf{0}, \mathbf{I}_J)$ verwendet. Für die Log-Likelihoodfunktion der *gesamten Daten* gilt dann

$$\mathcal{L}_{\text{Stufe 1}}(\boldsymbol{\theta}_{\text{Stufe 1}} | \mathbf{w}, \mathbf{x}, \mathbf{sd}) = \sum_{i=1}^I \ln(L_i(\mathbf{w}_i, \mathbf{x}, \mathbf{sd}_i, \boldsymbol{\delta}, \bar{\lambda}, \boldsymbol{\theta}_\Psi)). \quad (4.12)$$

Alle Parameter der ersten Stufe fasst $\boldsymbol{\theta}_{\text{Stufe 1}}$ zusammen und Schätzungen hierfür folgen aus der Maximierung von (4.12). Dieser *Maximum-Simulated-Likelihood (MSL)-Schätzer* besitzt die günstigen Eigenschaften der Konsistenz, asymptotischen Normalität und Effizienz, sofern hinreichend viele Züge zur Simulation von (4.11) eingesetzt werden (McFadden und Train 2000; Train 2009). Zwei Gründe sprechen dafür, dass diese Bedingung hier erfüllt ist:

1. Bei gleicher Anzahl der Züge führen Zufallszahlen, die auf Haltonsequenzen basieren, zu einer exakteren Lösung der Integrale als Pseudo-Zufallszahlen (Train 2009).

2. Die Schätzungen wurden ebenfalls mit $S = 1.000$ durchgeführt und die Ergebnisse bleiben stabil.

Es sei in Erinnerung gerufen, dass δ sämtliche Effekte auf Ebene der Durchschnittsnutzen in den Datendimensionen j und t beinhaltet. Hierzu gehören die zeitvariablen Markenwerte, die Marketing-Mix-Effekte, die UMC und die Effekte der CF. Diese Effekte können folglich nicht mehr fälschlicherweise durch individuelle Nutzenkomponenten aufgenommen werden und Ergebnisse hierzu sind auf der ersten Schätzstufe *unverzerrt*.

4.4.2 Zweite Stufe

Zur Schätzung der Populationsmittelwerte der Nutzenparameter (außer dem State Dependence-Effekt $\bar{\lambda}$, der schon auf der ersten Stufe ermittelt wurde) und den Effekten der CF (\mathbf{u}) kommen nun die geschätzten FE auf der zweiten Stufe als abhängige Variable zum Einsatz (Berry 1994; Berry *et al.* 2004; Chintagunta *et al.* 2005). Weil es sich nun nicht mehr um ein kompliziertes nichtlineares diskretes Wahlmodell handelt, ist es viel einfacher, Preisendogenität und Markenwertdynamik zu berücksichtigen. Hierzu wird die Durchschnittsnutzenfunktion (4.9) im Rahmen einer *gepoolten Regression* (der Dimension $J \cdot T$) mit *zeitvariablen Markenparametern* sowie *Marketing-Mix-Variablen* und *CF* als erklärenden Variablen verwendet (Kim 2010; Kim und Kim 2011).²⁰ Die UMC nehmen auf der zweiten Schätzstufe die Rolle des Störterms ein.

Zwei Gruppen von Parametern sind zu unterscheiden: 1.) Die zeitvariablen Markennutzen ($\bar{\alpha}_t$) und 2.) die konstanten Parameter auf Durchschnittsnutzenebene ($\theta_{\text{Stufe 2}}^N$). Zu letzteren zählen die Populationsmittelwerte (β), die CF-Effekte (\mathbf{u}) sowie die Standardabweichungen der UMC (σ_ε) und der dynamischen Markennutzen (σ_η). Zudem werden die Startwerte der zeitvariablen Markennutzen ($\bar{\alpha}_0$) als konstante Parameter mitgeschätzt. Damit Schätzungen für die CF vorliegen, gilt es ergänzend, die konstanten Parameter des Preismodells (4.8) zu ermitteln ($\theta_{\text{Stufe 2}}^P = [\gamma \ \sigma_\zeta]'$).

²⁰ Auf der zweiten Stufe könnte zusätzlich die *Schätzunsicherheit der FE* aus der ersten Stufe Berücksichtigung finden. Hierdurch verbessert sich die statistische Effizienz (Train 2009, S. 324 f.), allerdings existiert hierfür keine Standardsoftware, welche die KF-Schätzung zeitvariabler Parameter ermöglicht. Probeweise durchgeführte KF-Schätzungen mit selbst erstelltem Code zeigen, dass die resultierenden Unterschiede vernachlässigbar sind, sofern die geschätzten FE eine ähnlich hohe Präzision aufweisen, wie in den empirischen Studien in Abschnitt 4.5 (zwei- bis dreistellige t -Werte).

Auch auf der zweiten Stufe kommt ein likelihoodbasierter Schätzer zum Einsatz. Alle *konstanten* Modellparameter ($\boldsymbol{\theta}_{\text{Stufe 2}} = [\boldsymbol{\theta}_{\text{Stufe 2}}^N \ \boldsymbol{\theta}_{\text{Stufe 2}}^P]'$) werden gemeinsam geschätzt und die hierzu notwendigen Log-Likelihood-funktionen lauten (Kim und Kim 2011):

$$\mathcal{L}_{\text{Stufe 2}}^N(\boldsymbol{\theta}_{\text{Stufe 2}}^N | \boldsymbol{\delta}, \mathbf{x}, \boldsymbol{\zeta}^*) = \ln \left(\prod_{t=1}^T f(\boldsymbol{\delta}_t, \mathbf{x}_t, \boldsymbol{\zeta}_t^* | \boldsymbol{\theta}_{\text{Stufe 2}}^N) \right) \quad (4.13)$$

$$= -\frac{1}{2} \cdot \sum_{t=1}^T \left[\ln(2\pi^J |\boldsymbol{\Sigma}_t|) + \mathbf{e}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{e}_t \right] \text{ und} \quad (4.14)$$

$$\mathcal{L}_{\text{Stufe 2}}^P(\boldsymbol{\theta}_{\text{Stufe 2}}^P | \mathbf{p}, \mathbf{z}) = \ln \left(\prod_{t=1}^T \prod_{j=1}^J f(p_{jt}, z_{jt} | \boldsymbol{\theta}_{\text{Stufe 2}}^P) \right) \quad (4.15)$$

$$= -\frac{1}{2} \cdot \sum_{t=1}^T \sum_{j=1}^J \left[\ln(2\pi\sigma_\zeta^2) + (p_{jt} - \mathbf{z}_{jt}' \boldsymbol{\gamma})^2 / \sigma_\zeta^2 \right]. \quad (4.16)$$

Es sei daran erinnert, dass $\boldsymbol{\zeta}^* = (\mathbf{p} - \mathbf{z}\boldsymbol{\gamma})/\sigma_\zeta$ gilt. Wird dieser Term in (4.14) eingesetzt, folgt kompakt $\mathcal{L}_{\text{Stufe 2}}(\boldsymbol{\theta}_{\text{Stufe 2}} | \boldsymbol{\delta}, \mathbf{x}, \mathbf{p}, \mathbf{z}) = \mathcal{L}_{\text{Stufe 2}}^N + \mathcal{L}_{\text{Stufe 2}}^P$ und diese Funktion ist zur Schätzung von $\boldsymbol{\theta}_{\text{Stufe 2}}$ zu maximieren (Park und Gupta 2009).

Die Log-Likelihoodfunktion (4.16) basiert auf der Dichte der Normalverteilung des Störterms im Preismodells. Die Herleitung von (4.14) ist komplizierter. Wegen der Markenwertdynamik sind die FE *nicht intertemporal unabhängig* voneinander. Auf Basis der Prognosefehler (\mathbf{e}_t) und deren Kovarianz ($\boldsymbol{\Sigma}_t$) lässt sich die Dichte der multivariaten Normalverteilung für alle Zeitpunkte aber leicht darstellen (Harvey 1989, S. 126). Zur Herleitung dieser Terme ist der KF hilfreich.²¹ Dieser berechnet rekursiv die *ersten beiden Momente* der zeitvariablen Markenwerte (\mathbf{a} und \mathbf{P}) und es gilt für $t = 1, \dots, T$:

$$\mathbf{a}_{t|t-1} = \mathbf{a}_{t-1|t-1}, \quad \mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \mathbf{I}_J \sigma_\eta, \quad (4.17)$$

$$\mathbf{e}_t = \boldsymbol{\delta}_t - \mathbf{a}_{t|t-1} - \mathbf{x}_t \bar{\boldsymbol{\beta}} - (\mathbf{I}_J \boldsymbol{\zeta}_t^*) \boldsymbol{\iota}, \quad \boldsymbol{\Sigma}_t = \mathbf{P}_{t|t-1} + \mathbf{I}_J \sigma_\varepsilon, \quad (4.18)$$

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \boldsymbol{\Sigma}_t^{-1} \mathbf{e}_t, \quad \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \boldsymbol{\Sigma}_t^{-1} \mathbf{P}_{t|t-1}. \quad (4.19)$$

²¹ Das Vorgehen wird hier nur skizziert. Eine ausführliche Darstellung ist in Abschnitt 2.3 des 2. Kapitels der vorliegenden Arbeit bzw. bei Harvey (1989) oder Petris *et al.* (2009) zu finden.

Als Nebenprodukt des Algorithmus werden ebenfalls \mathbf{e}_t und $\mathbf{\Sigma}_t$ ermittelt. Ein Durchlauf des KF liefert also alle Werte zur Aufstellung von (4.14). Mit den geschätzten Startwerten $\mathbf{a}_{0|0} = \hat{\mathbf{a}}_0$ beginnt die KF-Rekursion. Die KF-Schätzung liefert hierdurch in Situationen ohne Dynamik ($\sigma_\eta = 0$) dieselben Resultate, wie eine Regression mit konstanten Parametern.

Nach der Schätzung aller konstanten Modellparameter lassen sich die zeitvariablen Markennutzen *glätten*. Hier werden die Schätzungen der ersten beiden Momente aus dem KF für $t = T - 1, T - 2, \dots, 1$ in den *Kalman Smoother* (KS) eingesetzt (Petrís *et al.* 2009, S. 61):

$$\mathbf{a}_{t|T} = \mathbf{a}_{t|t} + \mathbf{P}_{t|t} \mathbf{P}_{t+1|t}^{-1} (\mathbf{a}_{t+1|T} - \mathbf{a}_{t+1|t}) \text{ und} \quad (4.20)$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{P}_{t|t} \mathbf{P}_{t+1|t}^{-1} (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) (\mathbf{P}_{t|t} \mathbf{P}_{t+1|t}^{-1})'. \quad (4.21)$$

Die Schätzung der zeitvariablen Nutzen auf der zweiten Stufe hat mehrere günstige Eigenschaften (vgl. ausführlich Abschnitt 2.3 der vorliegenden Arbeit):

1. *Endogenität* ist durch CF zu handhaben, ohne dass es zu einer Beeinträchtigung der dynamischen Schätzung kommt (Kim und Kim 2011).
2. KF und KS können *nichtstationäre Dynamiken* (Random Walk-Prozesse) schätzen. Alternative Methoden benötigen restriktivere Annahmen (z. B. bzgl. zeitlicher Verläufe oder der Auswahl erklärender Variablen) und/oder Datentransformationen (z. B. Zeitdifferenzen), um Schätzungen hierfür zu ermöglichen, was eine direkte Interpretation der Ergebnisse erschwert (van Heerde *et al.* 2004).
3. Der KF kann problemlos *fehlende Werte* handhaben und liefert darüber hinaus Schätzungen für die Datenlücken, was in vielen Anwendungsfällen relevant ist (Petrís *et al.* 2009). Das gilt auch in den folgenden empirischen Studien.

4.5 Empirische Studien

Das zuvor präsentierte diskrete Wahlmodell mit zeitvariablen Markenwerten und das Vorgehen zur Schätzung dieses Modells wird nun für die beiden Produktkategorien *Margarine* und *Waschmittel* exemplarisch vorge-

stellt. Nach einer Beschreibung der Daten (Unterabschnitt 4.5.1) werden die Schätzergebnisse präsentiert (Unterabschnitt 4.5.2). Danach (Unterabschnitt 4.5.3) werden mögliche Implikationen untersucht, die sich aus den Markenwertergebnissen ableiten lassen.

4.5.1 Daten

Die *Daten* für beide Produktkategorien stammen aus unterschiedlichen Quellen. Deskriptive Statistiken hierzu sind in den Tabellen 4.4 (markenspezifische Variablen) und 4.5 (haushaltsspezifische Variablen) dargestellt.

Tabelle 4.4: Deskriptive Statistiken zu den markenspezifischen Variablen der Margarine- und Waschmitteldaten

Variable	Margarine			
	Blue Bonnet	I Cant't Believe It's Not Butter	Parkay	Shedd's
Marktanteil	0,291 (0,140)*	0,275 (0,130)	0,225 (0,112)	0,209 (0,125)
Preis	0,578 (0,096)	1,547 (0,071)	0,829 (0,085)	0,967 (0,063)
Feature	0,180 (0,322)	0,069 (0,173)	0,144 (0,271)	0,029 (0,145)
Display	0,033 (0,077)	0,000 (0,005)	0,022 (0,061)	0,000 (0,003)
Kosten	0,520 (0,049)	1,044 (0,020)	0,604 (0,031)	0,624 (0,020)

Variable	Waschmittel			
	Ariel	Denkmit	Persil	Spee
Marktanteil	0,179 (0,384)	0,563 (0,496)	0,115 (0,319)	0,143 (0,350)
Preis	4,322 (0,316)	2,408 (0,060)	4,820 (0,372)	3,123 (0,194)
Promotion	0,322 (0,472)	0,004 (0,067)	0,296 (0,458)	0,062 (0,242)

Anmerkung: (*) Mittelwert und Standardabweichung in Klammern. Die Variablen sind wie folgt skaliert: Preise sowie Kosten beziehen sich in der Margarinekategorie auf Packungen der Größe 16 oz und sind in \$(US) angegeben. In der Waschmittelkategorie handelt es sich um €-Preise pro Packung mit 20 Waschladungen. Die Promotionvariablen beziehen sich auf den Anteil der Wahlentscheidungen, die unter (den jeweiligen) Promotionbedingungen stattfinden. Bei den Marktanteilen handelt es sich um „Inside-Good Marktanteile“, sie sind demnach auf die Kategoriewahl bezogen. Die Marktanteile der Nichtwahloptionen betragen 90,0 % (Margarine) bzw. 89,2 % (Waschmittel) (vgl. hierzu auch Tabelle 4.5).

Für die *Margarinekategorie* handelt es sich um Daten aus der Gegend um Denver (USA), die einen Zeitraum von 117 Wochen abdecken (Anfang Januar 1993 bis Ende März 1995).²² Erfasst werden die Daten an den Kassen

²² Dieselben Daten wurden (u. a.) von Chintagunta *et al.* (2005), Dubé *et al.* (2009, 2010) sowie Conley *et al.* (2012) verwendet.

Tabelle 4.5: Deskriptive Statistiken zu den Haushalten und deren Einkaufsverhalten in der Margarine- und Waschmittelkategorie

	Margarine	Waschmittel
Wochen	105 (117)*	165 (226)
Einkäufe	56.138	70.185
Kategoriekäufe	5.693	7.569
Haushalte	992	2.055
Anteil SD-Käufe**	56,7%	71,1%

Anmerkung: (*) Der Wert in Klammern bezieht sich auf die gesamte Zeitspanne der zur Verfügung stehenden Daten. Es werden Wochen aus der Analyse entfernt, wenn weniger als drei Wahlentscheidungen für wenigstens eine der vier Marken existieren, weil andernfalls zu wenig Beobachtungen für eine reliable Schätzung der FE vorliegen (Chintagunta et al. 2005, S. 840). (**) Prozent der Wahlentscheidungen, bei denen dieselbe Marke gewählt wird wie bei dem letzten Kategoriekauf.

der Geschäfte durch den ScanTrack-Service von ACNielsen. Berücksichtigt werden Wahlentscheidungen für die vier Marken mit dem größten Marktanteil (*Blue Bonnet* (BB), *I Can't Believe It's Not Butter* (IC), *Parkay* (PA) und *Shedd's* (SH)) beim größten Händler²³ der Region und es wird nur die gängigste Packungsgröße von 16 oz (\approx 454 Gramm) betrachtet. Diese vier Marken zusammen decken mehr als 50 % der Gesamtnachfrage ab. 992 Haushalte werden in die Untersuchung einbezogen, welche insg. 56.138 Einkäufe tätigen, wovon 5.693 Margarinekäufe sind. Bei mehr als der Hälfte der Kategoriekäufe entscheiden sich die Haushalte für dieselbe Marke, wie beim letzten Kategoriekauf. Dies lässt darauf schließen, dass Kauffeedbackeffekte in der Margarinekategorie von Bedeutung sind (vgl. ausführlich Dubé et al. 2010).

Aus Tabelle 4.4 geht hervor, dass BB den höchsten (auf Kategoriekäufe bezogenen) Marktanteil besitzt, hiernach folgt direkt IC und mit etwas Abstand PA und SH. IC hat mit Abstand den höchsten Preis, allerdings mit einer eher geringen Variation im Vergleich zu BB und PA. Bei BB und PA werden häufiger *Promotionmaßnahmen* (insb. Features) eingesetzt. Sehr hilfreich bei den Margarinedaten ist die Verfügbarkeit der *Kostenvariable* (Großhandelspreise), die von Chintagunta et al. (2005) extra beschafft wurde, um Preisendogenität besser behandeln zu können.²⁴ Die Variable liefert aber auch interessante Einsichten bzgl. der Sortimentsstruktur und der Preissetzung des Händlers. BB ist eine niedrigpreisige Marke, mit wel-

²³ Der Name des Händlers ist nicht bekannt. Weitergehende Informationen zu den Daten sind Chintagunta et al. (2005) zu entnehmen.

²⁴ Conley et al. (2012) liefern Argumente für die Validität der Großhandelspreise als Instrumentalvariablen in diesem Margarinedatensatz.

cher der Händler mit durchschnittlich ca. 6 \$-Cent nahezu keinen Stückdeckungsbeitrag erzielt („loss-leader“). Durch häufige Preissenkungen und Promotionmaßnahmen werden aber regelmäßig Käufer zum Einkauf im betrachteten Geschäft bewegt. Der Stückdeckungsbeitrag der Premiummarke IC beträgt im Durchschnitt dagegen fast das Zehnfache pro Packung (ca. 50 \$-Cent), und auch die beiden anderen Marken PA und SH liegen mit ihren Stückdeckungsbeiträgen im zweistelligen \$-Cent-Bereich. Auf Basis dieser Angaben und der Tatsache, dass IC zu einem deutlich höheren Preis angeboten wird, ist zu schließen, dass IC einen hohen und BB einen niedrigen Markenwert besitzen dürfte. Die Daten lassen allerdings keine Vermutungen bzgl. der zeitlichen Markenwertentwicklung zu.

Die *Waschmitteldaten* stammen aus dem deutschen GfK-Haushaltspanel und erstrecken sich von Januar 2006 bis Mai 2010 (226 Wochen). Die Datenerfassung erfolgt mittels *Inhome-Scanning*²⁵ und es werden nur Einkäufe beim Drogeriemarkt *dm* betrachtet. Dieses Vorgehen hat zwei Gründe:

1. Deutschlandweit verwendet *dm* über alle Geschäfte einen sehr homogenen Marketing-Mix, was genau zur Modellspezifikation passt (vgl. Gleichung (4.1)) und Verzerrungen durch Datenaggregationen über die Geschäfte minimiert (Christen *et al.* 1997).
2. Es gibt in den Daten keine Informationen zu Promotionaktivitäten. Dies ist aber unkritisch, da *dm* vorrangig Preispromotions durchführt, die sich aus der charakteristischen Variation in den Preisdaten ableiten lassen.²⁶

Es werden außerdem nur kleine Packungen (ca. 20 Waschladungen) in die Analyse einbezogen und insg. gibt es mehr Beobachtungen als bei in der Margarinekategorie. 2.055 Haushalte sind für 70.185 Einkäufe verantwortlich und hiervon sind 7.569 Kategoriekäufe. Der Anteil der State Dependence-Käufe ist mit mehr als 70 % höher als in der Margarinekategorie, folglich spielt State Dependence bei Waschmitteln eine noch zentralere Rolle als bei Margarinen. In dieser Kategorie werden ebenfalls vier *Marken* untersucht, und zwar *Ariel* (AR), *Denkmit* (DM), *Persil* (PE) und *Spee* (SP), welche zusammen für mehr als 80 % des Waschmittelmarkts bei *dm* verantwortlich sind. Im Vergleich zur Margarinekategorie sind die

²⁵ Details zu der Datenquelle sowie eine Diskussion der Stärken und Schwächen unterschiedlicher Erfassungsmethoden liefern z. B. Günther *et al.* (2006).

²⁶ Promotions werden über temporäre Preisreduktionen identifiziert, die mehr als 10 % umfassen.

Marktanteile weniger einheitlich verteilt. Mit mehr als 50 % besitzt DM (die Handelsmarke von *dm*) eine sehr starke Marktposition. Die anderen drei Marken (AR, PE und SP) sind mit Marktanteilen zwischen 11 % und knapp 18 % deutlich kleiner. DM hat ferner den niedrigsten *Preis* und es werden fast keine *Promotionaktivitäten* durchgeführt. Für die hochpreisigen Marken AR und PE gibt es dagegen relativ häufig Preispromotions (jeweils ca. jede dritte Woche). SP liegt preislich zwischen AR und PE einerseits sowie DM andererseits und es gibt wenige Promotionaktivitäten. Für diese Kategorie liegen leider keine Kosteninformationen vor, weshalb wie von Villas-Boas und Winer (1999) vorgeschlagen, für jeden Zeitpunkt (t) *Preise früherer Zeitpunkte* ($t - 1$) als Instrument zur Behandlung des Endogenitätsproblems zum Einsatz kommen. Durch die Verwendung der exogenen Preisinformation aus der Vorperiode wird die Korrelation zwischen der Preisvariable und dem Störterm zum Zeitpunkt t „herausgetrennt“ (vgl. Gleichung (4.8)). Dieses Vorgehen ist in der Marketingliteratur nicht unüblich und ebenfalls z. B. bei Yang *et al.* (2003) oder Moon *et al.* (2006) zu finden.

Beide Datensätze besitzen hinreichend große Längsschnitt- (Wochen) und Querschnittdimensionen (Haushalte). Es findet eine enge Marktabgrenzung statt, mit Marken, die hinsichtlich ihrer technisch/funktionalen Eigenschaften gut substituierbar sind. Hierdurch wird sichergestellt, dass nach einer Kontrolle auf Effekte durch Marketing-Mix, State Dependence und UMC prinzipiell nur der Einfluss der Marke übrig bleibt. Beide Kategorien lassen sich als „reif“ bezeichnen, da sie seit Jahrzehnten existieren. Gleichzeitig gibt es ständig hohe Markeninvestitionen (z. B. in Form von TV-Werbung für Waschmittel) und es existieren interessante Marktstrukturen (z. B. mit sehr ausgeglichenen Marktanteilen in der Margarinekategorie, oder einer extrem starken Handelsmarke bei Waschmitteln). Demnach ist anzunehmen, dass deutliche Markenwertunterschiede über Marken und vor allem Zeitpunkte vorliegen. Es bleibt festzuhalten, dass sich beide Datensätze damit per se gut für eine Anwendung des postulierten Ansatzes zur Messung zeitvariabler Markenwerte eignen sollten.

4.5.2 Ergebnisse

Die Ergebnisse für die Parameterschätzungen werden getrennt nach *konstanten* und *zeitvariablen* Parametern präsentiert. Bevor die Schätzungen näher betrachtet werden, wird zunächst ein *Modellvergleich* anhand von *Gütekriterien* vorgenommen.

Modellvergleich

In Tabelle 4.6 sind Gütemaße (negativer \mathcal{L} -Wert sowie die Informationskriterien AIC und BIC) für die diskreten Wahlmodelle auf der ersten Schätzstufe abgetragen. Die Werte geben Aufschluss darüber, inwiefern *State Dependence*, *Heterogenität* und *FE* eine gewisse Erklärungskraft für die Markenwahl besitzen. Der negative \mathcal{L} -Wert verbessert (= verkleinert) sich mit der Zahl ins Modell aufgenommenen Parameter. AIC und BIC bestrafen jedoch zunehmende Modellkomplexität (Greene 2011, S. 533). In jeder Produktkategorie unterscheiden sich die Modelle durch drei Faktoren (State Dependence, Heterogenität und FE), weshalb es acht mögliche Modellvarianten gibt. Ausgehend vom einfachsten (MNL-)Modell werden zunehmend komplexere Spezifikationen untersucht. Die Identifizierbarkeit sämtlicher Modelle wird jeweils durch eine Eigenwertkontrolle der Hessematrix der maximierten Log-Likelihoodfunktionen überprüft (vgl. Haaijer *et al.* 1998).

Tabelle 4.6: Modellvergleich auf der ersten Stufe

Modelle*			Margarine			Waschmittel		
SD	Het.	FE	$-\mathcal{L}^{**}$	AIC***	BIC****	$-\mathcal{L}$	AIC	BIC
—	—	—	25813,3	51660,6	51723,1	32917,7	65847,4	65902,4
✓	—	—	24014,4	48044,8	48116,2	29683,0	59380,1	59444,2
—	✓	—	22334,7	44739,4	45052,1	27901,4	55856,8	56104,1
—	—	✓	25349,1	51546,2	55334,9	32515,8	66361,6	72452,3
✓	✓	—	22275,0	44637,9	45031,1	27619,7	55309,5	55630,0
✓	—	✓	23591,0	48031,9	51829,6	29301,2	59932,6	66023,2
—	✓	✓	21901,8	44713,5	48779,2	27483,8	56341,6	62633,7
✓	✓	✓	21851,6	44631,2	48777,8	27195,1	55780,2	62145,6

Anmerkung: (*) Das Basismodell (erste Zeile der Tabelle) ist ein MNL-Modell ohne State Dependence (SD), Heterogenität (Het.) und fixe Effekte (FE). (**) $-\mathcal{L}$ = negativer Wert der Log-Likelihoodfunktion an der Stelle ihres Maximums; (***) $AIC = -2 \cdot (\mathcal{L} - \# \text{ Parameter})$; (****) $BIC = -2 \cdot \mathcal{L} + \ln(\# \text{ Beobachtungen}) \cdot \# \text{ Parameter}$. Bei allen drei Gütemaßen weisen kleinere Werte auf eine höhere Güte hin. Fettgedruckte Werte kennzeichnen spaltenweise (also pro Gütemaß) das beste Modell.

Es ist für beide Kategorien ersichtlich, dass isoliert betrachtet sowohl die Berücksichtigung der State Dependence also auch der Heterogenität zu einer deutlichen Verbesserung des Datenfits führen, unabhängig davon, ob es FE in den Modellen gibt oder nicht. Bei einer gemeinsamen Berücksichtigung spielt State Dependence eine untergeordnete, aber noch immer relevante Rolle. Folglich ist eine gemeinsame Berücksichtigung zielführend, da andernfalls State Dependence einen Teil des Effekts der Heterogenität

aufnimmt und überschätzt wird (Keane 1997; Goldfarb 2006; Dubé *et al.* 2009, 2010). Die zusätzliche Aufnahme von FE in ein Modell führen in beiden Kategorien zu einer dreistelligen Reduktion der \mathcal{L} -Werte. Im Vergleich zu den Verbesserungen durch Heterogenität und State Dependence mag das auf den ersten Blick als ein wenig zufriedenstellendes Resultat erscheinen. Es sei aber in Erinnerung gerufen, dass Heterogenität und State Dependence „Phänomene“ auf der Individualebene darstellen und Haushaltspaneldaten hier besonders informativ sind (Dubé *et al.* 2009). FE erklären dagegen Variationen in den Wahlentscheidungen, die „nur“ in den Dimensionen Marke (j) und Zeit (t) stattfinden. Die große Anzahl zu schätzender FE (Margarine: $4 \cdot 105 = 420$; Waschmittel: $4 \cdot 165 = 660$) führt sogar dazu, dass es zu einer Verschlechterung der Datenfits gemäß BIC (Margarine) bzw. AIC und BIC (Waschmittel) kommt. Informationskriterien sind aber nicht das richtige Maß zur Modellselektion für die FE-Modelle, da sie vernachlässigen, dass die eigentliche Aufgabe der ersten Stufe eine Schätzung der FE ist (Chintagunta *et al.* 2005, S. 842). Schließlich ermöglichen die FE erst, Preisendogenität und vor allem dynamische Markennutzen auf der zweiten Stufe zu untersuchen.²⁷

Die *Ergebnisse* verdeutlichen die Notwendigkeit, State Dependence und Heterogenität auf der ersten Stufe gemeinsam adäquat zu berücksichtigen, da diese bereits einen Großteil der Variation in den Wahlraten erklären. Weil mit aggregierten Daten die Abbildung und Aufdeckung solcher Effekte schwierig ist, da keine individuellen Wahlmuster vorliegen, untermauern die Ergebnisse damit den Vorteil des im vorliegenden Beitrag beschriebenen Vorgehens, Haushaltspaneldaten, also Individualdaten, für eine dynamische Markenwertmessung nutzbar zu machen. Daher wird nachfolgend eine nähere Betrachtung der Modelle mit expliziter Berücksichtigung von State Dependence und Heterogenität der Konsumenten vorgenommen.

Konstante Parameter

Zunächst geht es um Schätzergebnisse der *konstanten* Parameter für beide Datensets. Der Fokus liegt auf drei Modellen, nämlich dem MXL-, dem MXL (CF)- und dem TVP-MXL (CF)-Modell²⁸, um entsprechend die Konsequenzen von *UMC*, *Endogenität* sowie *Markenwertdynamik* in diskreten

²⁷ Die Informationskriterien werden hier dennoch angegeben, da dies in der Literatur üblich ist und so die vorliegenden Befunde mit denen anderer Arbeiten verglichen werden können.

²⁸ „TVP“ steht für *time varying parameters*.

Wahlmodellen zu untersuchen und zu verstehen. Alle drei Modelle werden *zweistufig* geschätzt, wobei die erste Stufe das heterogene Logitmodell mit State Dependence und FE darstellt (letzte Zeile in Tabelle 4.6). Um ein besseres Verständnis für den FE-Ansatz zu entwickeln, wird zusätzlich das Ergebnis eines *einstufig* geschätzten MXL-Modells präsentiert, welches in der Marketingforschung den State of the Art darstellt und bspw. von Erdem *et al.* (2008) eingesetzt wird.

Aus Platzgründen zeigen die nachfolgenden Tabellen 4.7 und 4.8 nur einen Teil der Ergebnisse.²⁹ Sämtliche MXL-Spezifikationen besitzen eine flexible („volle“) Kovarianzmatrix (Ψ) zur Heterogenitätsmodellierung. Dargestellt werden aber lediglich die leichter interpretierbaren Standardabweichungen der heterogenen Variablen, anstatt der 36 (Margarine) bzw. 28 (Waschmittel) geschätzten Cholesky-Faktoren.³⁰

In Tabelle 4.7 sind die Ergebnisse für den *Margarinedatensatz* abgetragen. Viele Parameter sind auf dem 5 %-Niveau signifikant von Null verschieden. Erwartungsgemäß sind die Populationseffekte der Promotion- und State Dependence-Variablen positiv, die Preissensitivität dagegen ist negativ. Über alle (konstanten) Modelle hinweg besitzt IC den höchsten und BB den niedrigsten Markennutzen, die Marken PA und SH liegen dazwischen. Es ist zudem klar ersichtlich, dass die Haushalte sehr *heterogen in ihren Nutzenparametern* sind. Alles dies entspricht den Erwartungen.

Die zweistufig geschätzten Modelle lassen erkennen, dass *UMC* in den Daten eine Rolle spielen, da σ_ϵ signifikant positiv ist. Die Größenordnung des UMC-Effekts ist mit den Ergebnissen von Park und Gupta (2009) vergleichbar, die im Durchschnitt 0,538 ermitteln. Mit UMC im Modell fällt insb. die Heterogenität für die Markennutzen und die Preissensitivität geringer aus.

Die Resultate der beiden CF-Modelle zeigen, dass *Endogenität* in den *Margarinedaten* zu einer Unterschätzung der Preissensitivität führt. Die korrespondierenden mittleren Preiselastizitäten betragen z. B. für die zwei-

²⁹ Eine grafische Darstellung der 420 (Margarine) bzw. 660 (Waschmittel) geschätzten FE ist nachfolgend in Abbildung 4.2 zu finden. Sämtliche FE-Schätzungen fallen mit zwei- bis dreistelligen *t*-Werten sehr präzise aus. Die restlichen Schätzergebnisse sind im Anhang den Tabellen 4.10, 4.11, 4.12, und 4.13 zu entnehmen.

³⁰ Bei *k* heterogenen Parametern gibt es $k \cdot (k + 1) / 2$ Cholesky-Faktoren (Train 2009, S. 208 f.). Zur Berechnung der rücktransformierten Parameter und deren Standardfehler dient das simulationsbasierte Verfahren von Krinsky und Robb (1986, 1990) mit 10.000 Zufallszügen.

Tabelle 4.7: Schätzergebnisse für die konst. Parameter (Margarine)

Parameter	Eine Stufe	Zwei Stufen***		
	MXL	MXL	MXL (CF)	TVP-MXL (CF)
<i>Nutzenfkt.</i>				
$\bar{\alpha}_{BB}^*$	-2,239 (0,129)**	-2,246 (0,180)	-1,562 (0,531)	-1,078 (0,719)
$\bar{\alpha}_{IC}$	0,188 (0,334)	0,574 (0,451)	2,366 (1,356)	3,219 (1,885)
$\bar{\alpha}_{PA}$	-1,327 (0,179)	-1,336 (0,251)	-0,364 (0,743)	0,185 (1,024)
$\bar{\alpha}_{SH}$	-1,591 (0,215)	-1,388 (0,283)	-0,278 (0,847)	0,227 (1,178)
$\bar{\beta}_{Preis}$	-3,617 (0,213)	-3,871 (0,290)	-5,010 (0,872)	-5,599 (1,218)
$\bar{\beta}_{Feat.}$	0,293 (0,066)	0,311 (0,093)	0,187 (0,116)	0,089 (0,126)
$\bar{\beta}_{Disp.}$	1,361 (0,412)	1,401 (0,468)	1,007 (0,533)	0,286 (0,599)
$\bar{\lambda}$	0,401 (0,036)	0,365 (0,039)	0,365 (0,039)	0,365 (0,039)
<i>Heterogenität</i>				
$\sigma_{\nu_{\alpha BB}}$	2,263 (0,137)	1,948 (0,178)	1,948 (0,178)	1,948 (0,178)
$\sigma_{\nu_{\alpha IC}}$	3,087 (0,287)	3,032 (0,242)	3,032 (0,242)	3,032 (0,242)
$\sigma_{\nu_{\alpha PA}}$	2,220 (0,183)	1,436 (0,235)	1,436 (0,235)	1,436 (0,235)
$\sigma_{\nu_{\alpha SH}}$	2,216 (0,168)	2,079 (0,174)	2,079 (0,174)	2,079 (0,174)
$\sigma_{\nu_{\beta Preis}}$	1,848 (0,223)	1,477 (0,242)	1,477 (0,242)	1,477 (0,242)
$\sigma_{\nu_{\beta Feat.}}$	0,421 (0,093)	0,468 (0,080)	0,468 (0,080)	0,468 (0,080)
$\sigma_{\nu_{\beta Disp.}}$	1,520 (0,407)	2,153 (0,443)	2,153 (0,443)	2,153 (0,443)
$\sigma_{\nu_{\lambda}}$	0,576 (0,041)	0,349 (0,049)	0,349 (0,049)	0,349 (0,049)
<i>Störterm</i>				
σ_{ε}		0,442 (0,015)	0,423 (0,015)	0,395 (0,016)
<i>Endogenität</i>				
ι_{BB}			-0,015 (0,074)	0,074 (0,089)
ι_{IC}			0,301 (0,077)	0,302 (0,104)
ι_{PA}			-0,005 (0,070)	0,061 (0,095)
ι_{SH}			0,140 (0,085)	0,141 (0,104)
<i>Dynamik</i>				
σ_{η}				0,042 (0,015)
$\mathcal{L}_{Stufe 1}$	-22.274,950	-21.851,590	-21.851,590	-21.851,590
$\mathcal{L}_{Stufe 2}$		132,493	151,484	158,312
$R_{Stufe 2}^2$		0,608	0,642	0,706
$R_{Instrumente}^2$			0,962	0,962

Anmerkung: (*) Im TVP-MXL (CF)-Modell beziehen sich die $\bar{\alpha}_j$ -Parameter auf die Startwerte der zeitvariablen Markenutzen ($t = 0$). (**) Fettgedruckte Parameter sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden (zweiseitig) und die Standardfehler der Schätzungen werden in Klammern angegeben. (***) Alle zweistufigen Schätzungen basieren auf demselben MXL-Modell mit FE auf der ersten Stufe und folglich gibt es hier keine Unterschiede der Ergebnisse bzgl. Heterogenität und State Dependence. Die CF-Modelle werden inkl. des Preismodells geschätzt, die \mathcal{L} - und R^2 -Werte der zweiten Stufe beziehen sich für Vergleichszwecke jedoch nur auf das Nachfragemodell.

stufigen Modelle: $-3,701$ (MXL), $-4,790$ (MXL (CF)) und $-5,353$ (TVP-MXL (CF)).³¹ Diese Elastizitätswerte sind (betragsmäßig) recht hoch (Bijmolt *et al.* 2005), passen aber z. B. zu den Ergebnissen von Chintagunta *et al.* (2005). Die Richtung der Endogenitätsverzerrung wurde bereits in vielen anderen Produktkategorien nachgewiesen (z. B. bei Ketchup von Villas-Boas und Winer 1999). Eine Umrechnung der ι_j -Parameter in Korrelationen (zwischen Preis- und Nutzenschocks) erleichtert deren Interpretation. Es gilt $\rho = \tau / \sqrt{1 + \tau^2}$, mit $\tau = \iota / \sigma_\epsilon$ und hiermit folgen z. B. für IC und SH Werte von 0,580 bzw. 0,314 für das MXL (CF)-Modell. Korrelationen in dieser Größenordnung interpretieren Jiang *et al.* 2009 als „moderat“.

Signifikant positiv ist σ_η im TVP-MXL (CF)-Modell, wodurch *Markenwertdynamik* in den Margarinaten belegt wird. Das TVP-MXL (CF)-Modell ist daher dem MXL (CF)-Modell (statistisch) überlegen. Es fällt zudem auf, dass die Berücksichtigung zeitvariabler Markennutzen einerseits die Preissensitivität erhöht, andererseits die Populationseffekte der Promotionvariablen senkt. Dies mag daran liegen, dass Marketing-Mix-Effekte in statischen Modellen wenigstens einen Teil der Dynamik aufnehmen (Harvey 1989, S. 365 ff.). Der geschätzte Wert von σ_η erscheint zunächst recht klein und liegt auch unter den Werten von Lachaab *et al.* (2006) oder Hernández-Mireles und Fok (2011). Diese Beiträge modellieren allerdings monatliche Markenwertschwankungen. Außerdem wären auch viel größere Werte für σ_η unrealistisch, da es sich um die Standardabweichung der Störterme von Random Walk-Prozessen handelt und davon auszugehen ist, dass sich Markenwerte eher „langsam“ entwickeln (Ailawadi *et al.* 2003).

Die R^2 -Werte zeigen, dass die Modelle auf der zweiten Stufe (also auf Ebene der FE) eine *gute Datenanpassung* besitzen. Vor allem die Varianzaufklärung von 70 % des TVP-MXL (CF)-Modells ist zufriedenstellend. Der R^2 -Wert von 0,962 der Preisregression zeigt, dass die herangezogenen Instrumentalvariablen eine sehr große Erklärungskraft besitzen und das Problem schwacher Instrumente daher keine Rolle spielt.

Die Ergebnisse in Tabelle 4.8 für die *Waschmittel*kategorie weisen allgemein eine hohe Ähnlichkeit zu denen der Margarinekategorie auf. Als zentrale Gemeinsamkeiten sind festzuhalten, dass *Marketing-Mix-Effekte* plausible Vorzeichen besitzen und *Heterogenität*, *Endogenität* sowie *zeitvariable Markennutzen* eine wichtige Rolle spielen. Zudem sinkt auch die Heterogenität

³¹ Die Tatsache, dass nur ι_{IC} einen signifikanten Effekt aufweist, ist im Rahmen des CF-Ansatzes unproblematisch (Pancras 2010). Ferner belegen Likelihood-Quotienten-Tests die gemeinsame Signifikanz der vier CF-Parameter in beiden CF-Modellen.

Tabelle 4.8: Schätzergebnisse für die konst. Parameter (Waschmittel)

Parameter	Eine Stufe		Zwei Stufen***	
	MXL	MXL	MXL (CF)	TVP-MXL (CF)
<i>Nutzenfkt.</i>				
$\bar{\alpha}_{AR}^*$	-1,883 (0,344)**	-2,476 (0,293)	-2,329 (0,405)	-1,827 (0,489)
$\bar{\alpha}_{DM}$	-2,093 (0,190)	-2,485 (0,162)	-2,405 (0,226)	-2,135 (0,296)
$\bar{\alpha}_{PE}$	-1,803 (0,372)	-2,758 (0,324)	-2,591 (0,449)	-1,877 (0,539)
$\bar{\alpha}_{SP}$	-3,328 (0,252)	-3,919 (0,209)	-3,812 (0,292)	-3,255 (0,370)
$\bar{\beta}_{Preis}$	-0,745 (0,078)	-0,616 (0,066)	-0,650 (0,093)	-0,778 (0,108)
$\bar{\beta}_{Promo}$	0,151 (0,067)	0,255 (0,050)	0,237 (0,051)	0,193 (0,035)
$\bar{\lambda}$	0,954 (0,033)	0,937 (0,036)	0,937 (0,036)	0,937 (0,036)
<i>Heterogenität</i>				
$\sigma_{\nu_{\alpha AR}}$	1,100 (0,354)	0,846 (0,490)	0,846 (0,490)	0,846 (0,490)
$\sigma_{\nu_{\alpha DM}}$	2,140 (0,126)	1,649 (0,105)	1,649 (0,105)	1,649 (0,105)
$\sigma_{\nu_{\alpha PE}}$	2,312 (0,335)	1,676 (0,293)	1,676 (0,293)	1,676 (0,293)
$\sigma_{\nu_{\alpha SP}}$	2,597 (0,171)	2,392 (0,225)	2,392 (0,225)	2,392 (0,225)
$\sigma_{\nu_{\beta Preis}}$	0,541 (0,072)	0,319 (0,044)	0,319 (0,044)	0,319 (0,044)
$\sigma_{\nu_{\beta Promo}}$	0,326 (0,063)	0,595 (0,085)	0,595 (0,085)	0,595 (0,085)
$\sigma_{\nu_{\lambda}}$	0,855 (0,044)	0,927 (0,043)	0,927 (0,043)	0,927 (0,043)
<i>Störterm</i>				
σ_{ε}		0,430 (0,012)	0,423 (0,012)	0,397 (0,012)
<i>Endogenität</i>				
ι_{AR}			0,102 (0,037)	0,104 (0,035)
ι_{DM}			0,111 (0,159)	0,121 (0,162)
ι_{PE}			-0,029 (0,029)	-0,003 (0,029)
ι_{SP}			-0,055 (0,063)	-0,032 (0,060)
<i>Dynamik</i>				
σ_{η}				0,034 (0,009)
$\mathcal{L}_{Stufe 1}$	-27.619,730	-27.195,090	-27.195,090	-27.195,090
$\mathcal{L}_{Stufe 2}$		230,771	237,775	247,018
$R^2_{Stufe 2}$		0,756	0,762	0,800
$R^2_{Instrumente}$			0,963	0,963

Anmerkung: (*) Im TVP-MXL (CF)-Modell beziehen sich die $\bar{\alpha}_j$ -Parameter auf die Startwerte der zeitvariablen Markennutzen ($t = 0$). (**) Fettgedruckte Parameter sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden (zweiseitig) und die Standardfehler der Schätzungen werden in Klammern angegeben. (***) Alle zweistufigen Schätzungen basieren auf demselben MXL-Modell mit FE auf der ersten Stufe und folglich gibt es hier keine Unterschiede der Ergebnisse bzgl. Heterogenität und State Dependence. Die CF-Modelle werden inkl. des Preismodells geschätzt, die \mathcal{L} - und R^2 -Werte der zweiten Stufe beziehen sich für Vergleichszwecke jedoch nur auf das Nachfragemodell.

in den Markennutzen und dem Preisparameter bei einem zweistufigen Vorgehen (Chintagunta *et al.* 2005).

Die CF-Parameter von AR und DM sind deutlich positiv, allerdings ist nur ι_{AR} signifikant von Null verschieden.³² Die Preiselastizitäten von $-2,207$ (MXL), $-2,329$ (MXL (CF)) und $-2,788$ (TVP-MXL (CF)) sowie die Richtung der Endogenitätsverzerrung passen in dieser Kategorie sehr gut zu typischen Resultaten in der Marketingliteratur (Bijmolt *et al.* 2005). Es ergeben sich nun Korrelationen von bspw. $0,234$ für AR und $0,254$ für DM und diese Werte liegen unter denen der Margarinekategorie. Dafür wirkt sich diesmal die Berücksichtigung der Dynamik in den Markennutzen stärker auf den (betragsmäßigen) Anstieg der Preiselastizität aus.

Die Ergebnisse der Waschmittelkategorie weisen aber auch einige erwähnenswerte *Besonderheiten* auf. Der State Dependence-Effekt fällt bei Waschmitteln besonders hoch aus, was dazu passt, dass es in dieser Kategorie einen sehr großen Anteil an aufeinanderfolgenden Käufen derselben Marken gibt (vgl. Tabelle 4.5). Die Ordnung der *Markennutzen* ist nun weniger eindeutig als zuvor, was die Wichtigkeit einer flexiblen Modellspezifikation betont.

Der geschätzte σ_ε -Wert passt auch zu den Ergebnissen von Park und Gupta (2009), hier ergeben sich aber weitreichendere Konsequenzen als bei den Margarinedaten. Dieser Punkt ist vor allem aus dem Vergleich der Mittelwerte der Nutzenparameter in der ein- vs. zweistufigen Schätzung des MXL-Modells ersichtlich, die sich durch die Berücksichtigung von UMC deutlich unterscheiden. Wieder zeigt sich die Notwendigkeit einer Modellspezifikation, die wichtige empirische Phänomene (z. B. UMC) abbildet. Die Datenanpassung auf der zweiten Schätzstufe fällt mit Varianzaufklärungen von mehr als 75 % noch besser aus als in der Margarinekategorie.

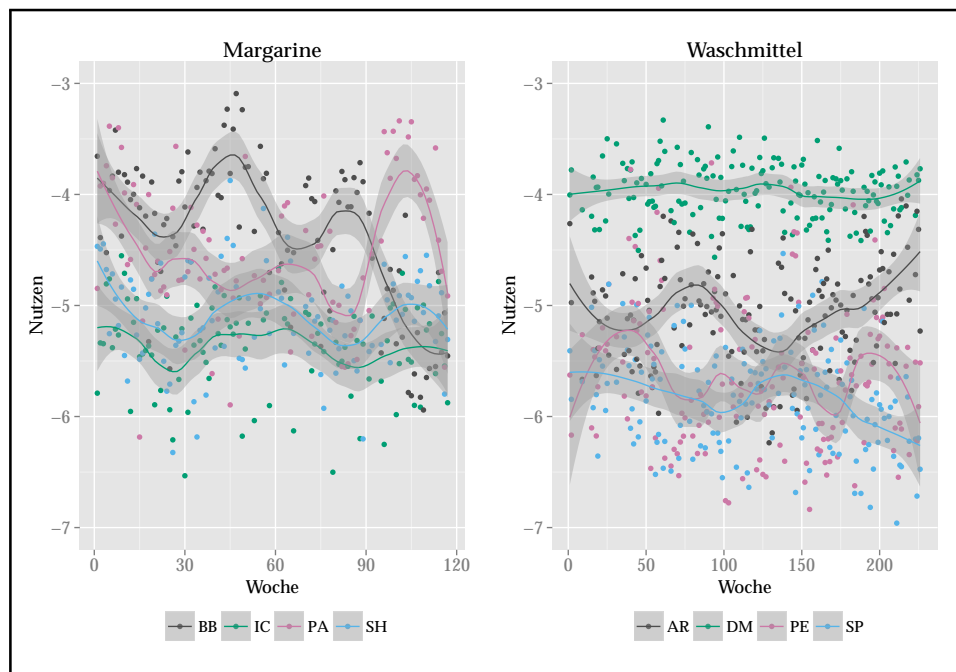
Insgesamt lässt sich festhalten, dass es für viele Modellergebnisse eine Rolle spielt, welche Spezifikation zu Grunde liegt. Endogenität führt in beiden Kategorien zu einer Unterschätzung der Preiseffekte. Darüber hinaus führt ebenfalls die Vernachlässigung zeitvariabler Markennutzen zu unterschätzten Preiseffekten, was ein interessantes und neues empirisches Ergebnis darstellt. UMC sind ferner wichtig, damit die Schätzung der Heterogenität nicht zu hoch ausfällt.

³² Auch in der Waschmittelkategorie zeigen Likelihood-Quotienten-Tests, dass die vier CF-Parameter in beiden Modellen mit CF gemeinsam signifikant sind.

Zeitvariable Parameter

Die Schätzergebnisse in den Tabellen 4.7 und 4.8 haben bereits gezeigt, dass das TVP-MXL(CF)-Modell den besten *Datenfit* aufweist und die Vorzeichen der konstanten Parameter den Erwartungen entsprechen. Bevor die *zeitvariablen Parameter* dieses Modells näher untersucht werden, ist eine *visuelle Inspektion der FE* informativ (Gelman und Hill 2007, S. 551 ff.). Dieser Blick auf die zeitliche Verteilung der latenten FE soll Indizien dafür liefern, ob Dynamik in den Markenwerten vorliegt.

Abbildung 4.2: Zeitliche Verteilung der Werte für die fixen Effekte im Margarine- und Waschmittelmarkt

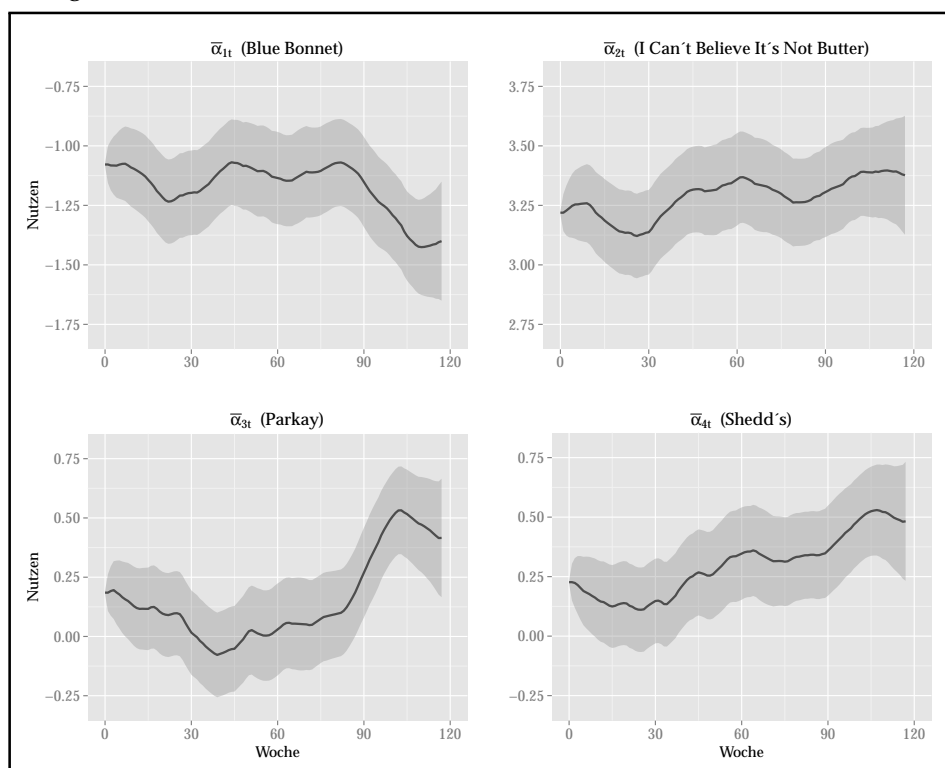


Anmerkung: Die geschätzten FE sind als Punkte dargestellt. Bei den Linien handelt es sich um Glättungen der Punkte mithilfe lokal gewichteter Regressionen (LOESS). Die 95 %-Konfidenzintervalle sind dunkelgrau hinterlegt eingezeichnet.

Abbildung 4.2 zeigt die zeitlichen Verläufe der FE beider Märkte. Die negativen FE-Werte lassen sich zunächst durch die hohen Marktanteile der Nichtwahloption von ca. 90 % in beiden Kategorien begründen (vgl. Tabelle 4.5). Die FE unterscheiden sich deutlich über Marken und es gibt starke zeitliche Schwankungen. Ersteres lässt darauf schließen, dass die Marken nicht denselben Nutzen stiften. Marketing-Mix-Effekte auf Populations-ebene und UMC erklären dagegen das zeitliche „Rauschen“. Änderungen im Markenwert sind aber eher mittel- oder langfristiger Natur (Ailawadi

et al. 2003), und ob diese vorliegen, lässt sich durch die ebenfalls in Abbildung 4.2 dargestellten *LOESS*-Glättungen („lokal gewichtete Regression“, vgl. Cleveland 1979 oder Fahrmeir *et al.* 2013) ergründen.³³ Solche mittel- und langfristigen Trends sind in beiden Märkten zu erkennen, obgleich sie für einzelne Marken unterschiedlich stark ausfallen. Die FE-Glättungen von DM im Waschmittelmarkt verlaufen sehr stabil, dagegen zeigt BB im Margarinemarkt klare Veränderungen.

Abbildung 4.3: Grafische Darstellung der zeitlichen Markenwertverläufe (Margarine)



Anmerkung: Die Markenwerte sind in Nutzeneinheiten skaliert. Zur besseren Vergleichbarkeit besitzt jede Grafik einen Wertebereich von einer Nutzeneinheit. Somit können absolute Werte von Grafik zu Grafik unterschiedlich ausfallen, ein Vergleich der Variation der Werte wird hierdurch aber erleichtert. Es sei zudem daran erinnert, dass lediglich die Differenzen der $\bar{\alpha}_{jt}$ -Parameter interpretierbar sind und nicht die absoluten Werte. Die 95%-Konfidenzintervalle sind dunkelgrau hinterlegt eingezeichnet.

Abbildung 4.3 zeigt nun die *KS*-Ergebnisse für die zeitvariablen Markenwerte in der Margarinekategorie. Diese Trajektorien stellen das *Kernergebnis*

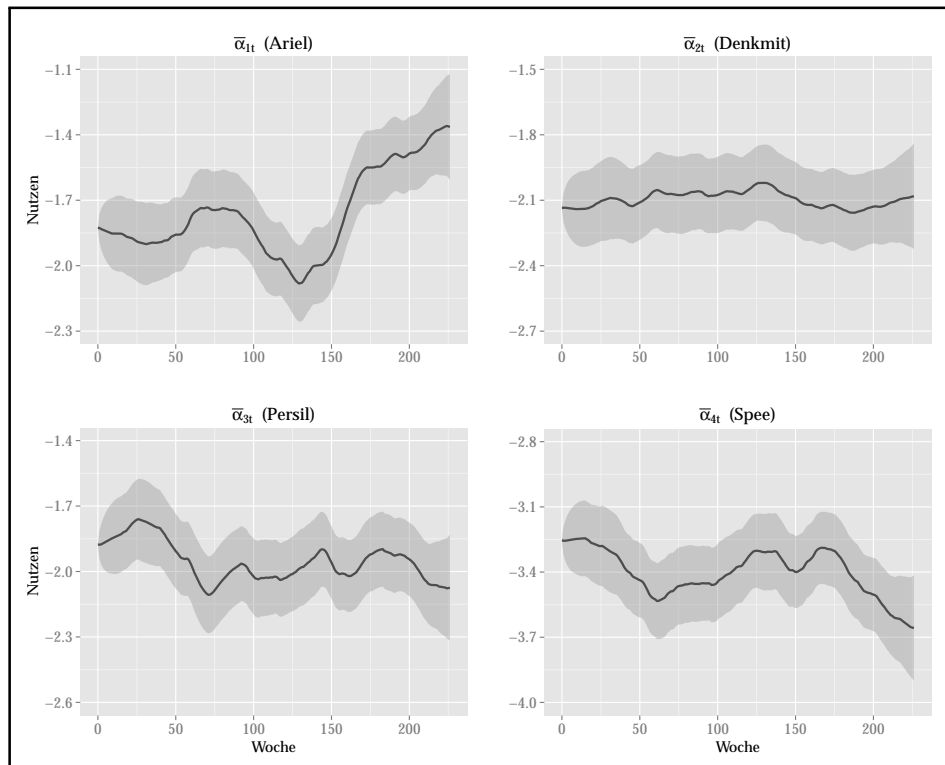
³³ Analog schlägt Albers (2012) *gleitende Durchschnitte* zur Bestimmung der (nichtlinearen) Funktionsformen in der Nachfragemodellierung vor.

der gesamten Untersuchung dar. Alle vier Marken weisen schwankende Markenwerte auf und diese sind fast alle über den gesamten Zeitraum signifikant von Null verschieden.³⁴ Die Entwicklungen sind jedoch sehr unterschiedlich, unabhängig voneinander und nicht monoton. Hier zeigt sich ein Vorteil der KF/KS-Schätzung, die lediglich *einen* Varianzparameter benötigt, um diese unterschiedlichen Dynamiken zugänglich zu machen. Ein weiterer Vorteil ist die Überwindung fehlender Werte mithilfe der KF/KS-Schätzung. In der Margarinekategorie fehlen mehr als 10 % der wöchentlichen Beobachtungen und trotzdem ist diese Methode in der Lage, Schätzungen für alle Wochen zu erzielen. Die generellen Verläufe sind auch nicht zu volatil; sie passen daher gut zum dynamischen Markenwertverständnis von Ailawadi *et al.* (2003). BB zeigt die ersten 90 Wochen einen recht stabilen Markenwert, der dann allerdings abnimmt. Der Markenwert von IC weist zwar Schwankungen auf, eine Betrachtung des 95 %-Konfidenzintervalls zeigt jedoch, dass diese nicht signifikant sind, da über den gesamten Zeitraum waagerechte Linien (hier z. B. mit dem Wert 3,25) das Konfidenzintervall nicht verlassen. PA verliert im ersten Jahr an Markenwert, dieser erholt sich danach aber wieder und erfährt eine deutliche Steigerung in den letzten Wochen des Untersuchungszeitraums. SH weist einen positiven Trend im Markenwert auf. Die absoluten Markenwerte im Vergleich bestätigen die Annahmen, dass IC (BB) generell einen hohen (niedrigen) Markenwert besitzt.

Die *zeitvariablen Markenwerte* der *Waschmittelkategorie* stellt Abbildung 4.4 dar. Auffällig ist hier zunächst, dass der Markenwert von DM stabil bleibt. Das ist eine interessante Erkenntnis und passt zu der Intuition, dass Handelsmarken weniger starken Schwankungen durch Trends und Markeninvestitionen unterliegen. Die beiden Marken PE und SP weisen beide moderate zeitliche Variation auf. Insbesondere in den letzten 2 Jahren bleibt der Markenwert von PE stabil; SP hat eine recht konstante Phase in der Mitte des Untersuchungszeitraums. Die stärksten Markenwertdynamiken gibt es bei AR. Nach einem Rückgang des Markenwerts ab Woche 90, erholt sich die Marke ein Jahr später und kann bis zum Ende des Datenfensters weiter deutlich zulegen. Diese Entwicklung bei AR ist insb. interessant im Vergleich zu dem Hauptkonkurrenten PE. Am Tiefpunkt von AR waren beide Marken bzgl. ihres Markenwerts ungefähr auf gleicher Höhe. Die positive Entwicklung von AR sollte dem Management von PE aber zu denken geben. Auch wenn die eigene Markenwertposition recht stabil

³⁴ Ausnahmen stellen die Markenwerte von PA (in den Wochen 10 bis 80) und SH (in den Wochen 10 bis 40) dar.

Abbildung 4.4: Grafische Darstellung der zeitlichen Markenwertverläufe (Waschmittel)



Anmerkung: Die Markenwerte sind in Nutzeneinheiten skaliert. Zur besseren Vergleichbarkeit besitzt jede Grafik einen Wertebereich von 1,2 Nutzeneinheiten. Somit können absolute Werte von Grafik zu Grafik unterschiedlich ausfallen, ein Vergleich der Variation der Werte wird hierdurch aber erleichtert. Es sei zudem daran erinnert, dass lediglich die Differenzen der $\bar{\alpha}_{jt}$ -Parameter interpretierbar sind und nicht die absoluten Werte. Die 95%-Konfidenzintervalle sind dunkelgrau hinterlegt eingezeichnet.

bleibt, offenbart der Blick zu AR klaren Handlungsbedarf. DM nimmt eine Sonderstellung im Vergleich der vier Marken ein. Zum einen zeigt sich, dass der Markenwert niedriger ausfällt, als die sehr starke Marktposition mit mehr als 50 % Marktanteil vermuten lässt. Andererseits besitzt DM nicht den niedrigsten Markenwert (sondern SP). DM befindet sich eher auf einem ähnlichen Markenwertniveau wie PE. Dies zeigt, dass auch Handelsmarken erhebliche Markenwerte aufweisen können (vgl. Sattler 1998).

Bevor auf weitere Implikationen eingegangen wird, ist eine *Validierung* des zeitvariablen Markennutzens als Markenwertmaß angebracht. Hierzu werden Korrelationen mit alternativen Maßen aus der Literatur berechnet (vgl. Ailawadi *et al.* 2003). Hernández-Mireles und Fok (2011) verwenden so-

wohl psychische Größen (z. B. „Top of Mind-Awareness“) als auch Marktgrößen (Marktanteil). Sriram *et al.* (2007) nutzen dagegen nur Marktgrößen (Absatz-, Preis- und Umsatzpremium). Weil Daten zu psychischen Größen für beide Kategorien nicht vorhanden sind, ist hieran ein Vergleich nicht durchführbar. Zudem sind nur für die Margarinekategorie „echte“ Marktdaten verfügbar. In der Waschmittelkategorie erfolgt aus diesem Grund eine Aggregation der Haushaltspaneldaten. Für die Waschmittelkategorie werden die Premiummetriken mit Bezug zur Handelsmarke DM berechnet. Weil es in der Margarinekategorie keine Handelsmarke gibt, wird BB, die günstigste Marke mit dem geringsten Nutzen, als Referenzmarke für die Berechnung der Premiummetriken herangezogen.

Die durchschnittlichen Korrelationen³⁵ zwischen dem Markennutzen und dem Marktanteil, Preis-, Absatz- und Umsatzpremium lauten in der Margarinekategorie jeweils 0,368, -0,505, 0,527 und 0,499. Die entsprechenden Werte für Waschmittel betragen 0,385, 0,125, 0,297 und 0,498. Sämtliche Korrelationen sind statistisch signifikant. Ungewöhnlich ist lediglich der negative Korrelationswert zwischen dem Markennutzen und dem Preispremium in der Margarinekategorie. Dieser „Ausreißer“ ist aber auf den überproportionalen Anstieg des Preises von BB im Vergleich zu den Konkurrenten zurückzuführen. Die restlichen Korrelationen entsprechen ungefähr denen bei Sriram *et al.* (2007) sowie Hernández-Mireles und Fok (2011) und belegen empirisch, dass es sich beim dynamischen Markennutzen um ein valides Markenwertmaß handelt. Deutlich höhere Korrelationen wären vor dem Hintergrund der in der Einleitung formulierten Kritik an Marktgrößen, welche neben Markeneffekten viele Ursachen für Variation besitzen, auch nicht zwingend wünschenswert.

4.5.3 Implikationen

Die Markenwerte lassen sich im Rahmen der Modellauswertung gut interpretieren und es ist durchaus vorteilhaft, dass die Ergebnisse auf Nutzebene vorliegen. Dies ermöglicht nämlich eine direkte Verbindung zum Wahlverhalten im Logitmodell. Für Managementimplikationen eignen sich

³⁵ Die Korrelationen werden pro Maß sowie Marke berechnet und dann über Marken gemittelt. Hierdurch wird sichergestellt, dass nur gemeinsame Variation *über die Zeit* in die Korrelation einfließt. Es findet also eine Validierung mit Bezug zur *dynamischen Markenwertmessung* statt, was vor dem Hintergrund des Untersuchungsgegenstands des vorliegenden Beitrags sinnvoll erscheint. Gepoolte Korrelationen inkl. der Variation im Querschnitt über Marken fallen noch höher aus.

aber eher *monetäre Größen*, weil diese mehr Entscheidungsrelevanz besitzen. Ferner erleichtern sie die Vergleichbarkeit über Datensätze (z. B. verschiedene Händler, Länder usw.) hinweg. Aus diesem Grund werden nachfolgend mehrere monetäre Größen berechnet, die sich aus den Modellergebnissen ableiten lassen. Zuerst werden *Zahlungsbereitschaften* (ZB) für Marken (monetäres Äquivalent des Markennutzens) errechnet und untersucht. Diese werden zusätzlich mit *Wechselkosten* (monetäres Äquivalent der State Dependence, Dubé *et al.* 2009, 2010) verglichen. Schließlich erfolgt beispielhaft eine Simulation des Einflusses des Markenwerts auf die *Gesamtmarktnachfrage* und *-umsätze*.

Zahlungsbereitschaften für Marken

ZB in Nutzenmodellen mit linear spezifiziertem Preis folgen aus der Division eines Nutzenunterschieds durch den Absolutwert des Preisparameters³⁶ (Sonnier *et al.* 2007; Train 2009). Die ZB gibt demnach an, wie hoch der Preis sein muss, um einen spezifischen Nutzenunterschied (hier bedingt durch Marken) auszugleichen.

Weil es Heterogenität in den Parametern gibt, sind auch die ZB heterogen. Aufgrund der Normalverteilungsannahme³⁷ kommt es nun aber zu einem Problem. Mit einer positiven Wahrscheinlichkeit existieren Personen in der Population, die eine Preissensitivität von Null besitzen. Hieraus folgt, dass die Erwartungswerte der ZB *nicht definiert* sind. Ein pragmatischer Ausweg aus diesem „Dilemma“ ist die Verwendung des *Medians* (anstatt des Mittelwerts) als Lageparameter für die ZB (Bliemer und Rose 2013). Die Wahl des Medians lässt sich auch damit begründen, dass hiermit *realistischere* Werte für ZB resultieren, da Extremwerte weniger stark ins Gewicht fallen (Sonnier *et al.* 2007).

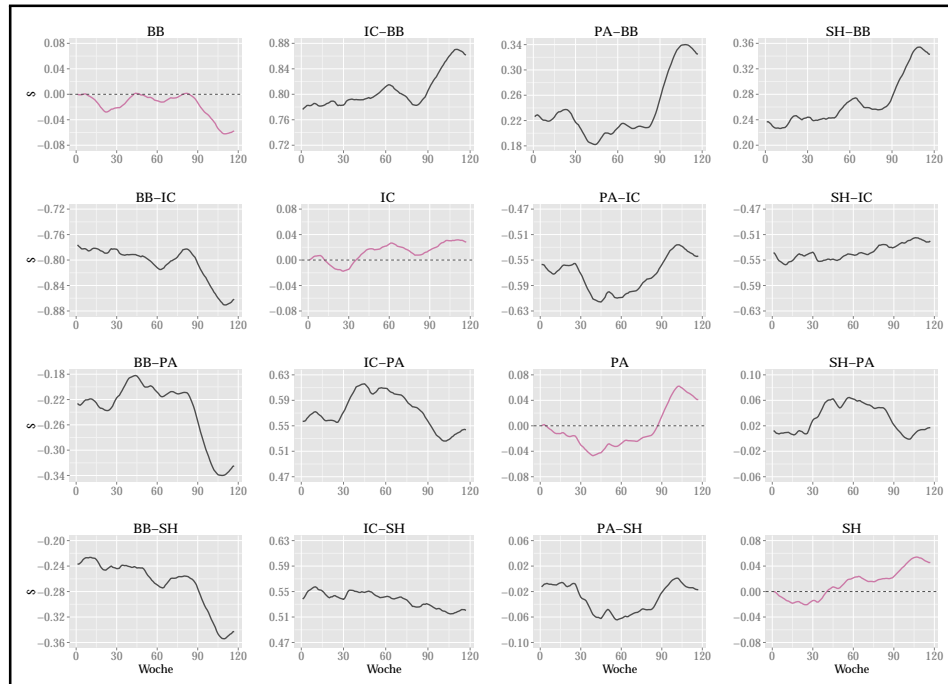
Die Berechnungen der Marken-ZB erfolgt mithilfe des Simulationsverfahrens von Krinsky und Robb (1986, 1990). Weil die Mittelwerte der Markennutzen wöchentlich variieren, wird die Simulation *für jede Woche* einzeln durchgeführt. Es werden jeweils $S = 10.000$ Züge aus den Heterogenitätsverteilungen der Parameter gezogen (vgl. Gleichung (4.2)). Für jeden Zug s zum Zeitpunkt t wird zuerst $\hat{\alpha}_{sjt}$ (Markennutzendifferenz zwischen zwei betrachteten Marken) oder $\check{\alpha}_{sjt}$ (Markennutzendifferenz einer Marke

³⁶ An dieser Stelle zeigt sich, warum die Behandlung der Preisendogenität wichtig ist. Wird nämlich die Preissensitivität *unterschätzt*, was in den empirischen Studien der Fall gewesen wäre (vgl. die Tabellen 4.7 und 4.8), würden sämtliche ZB *überschätzt*.

³⁷ Die Annahme war notwendig für die Trennung in zwei Schätzstufen, welche die Behandlung der Endogenität und Dynamik erst erlaubt.

bezogen auf $t = 1$) berechnet und dann durch $|\beta_{\text{Preis}}|$ dividiert. Danach wird der Median pro Woche über alle S Züge berechnet.

Abbildung 4.5: Grafische Darstellung der zeitlichen Zahlungsbereitschaftsverläufe (Margarine)



Anmerkung: Alle Zahlungsbereitschaften sind in \$ skaliert und gelten pro Packung. Zur besseren Vergleichbarkeit besitzt jede Grafik einen Wertebereich von 0,16\$. Somit können absolute Werte von Grafik zu Grafik unterschiedlich ausfallen, ein Vergleich der Variation der Werte ist hierdurch aber erleichtert.

In Abbildung 4.5 sind die ZB für die Margarinemarken abgetragen. Auf der Diagonalen sind die ZB pro Marke im Vergleich zur ersten Woche abgetragen und es ist zu erkennen, dass ihre Verläufe zu den Markennutzenentwicklungen in Abbildung 4.3 passen. Das war wegen der hohen Preissensitivität und mäßigen Heterogenität auch zu erwarten, denn hierdurch wird der Nutzen „nahezu linear“ durchskaliert.³⁸ Monetär ausgedrückt liegen die Entwicklungen der ZB im \$-Cent-Bereich (ca. 0,05\$ bis 0,10\$). Das mag nicht als viel erscheinen, im Vergleich zum durchschnittlichen Kategoriepreis³⁹ von ca. 0,967\$ betreffen diese zeitlichen Änderungen aber

³⁸ Man bedenke aber, dass dies nicht zwingend der Fall sein muss und die Simulation demnach unverzichtbar ist, um eine verlässliche Schätzung der ZB zu erhalten.

³⁹ Der Kategoriepreis wird als marktanteilsgewichteter Durchschnittspreis über die Marken berechnet.

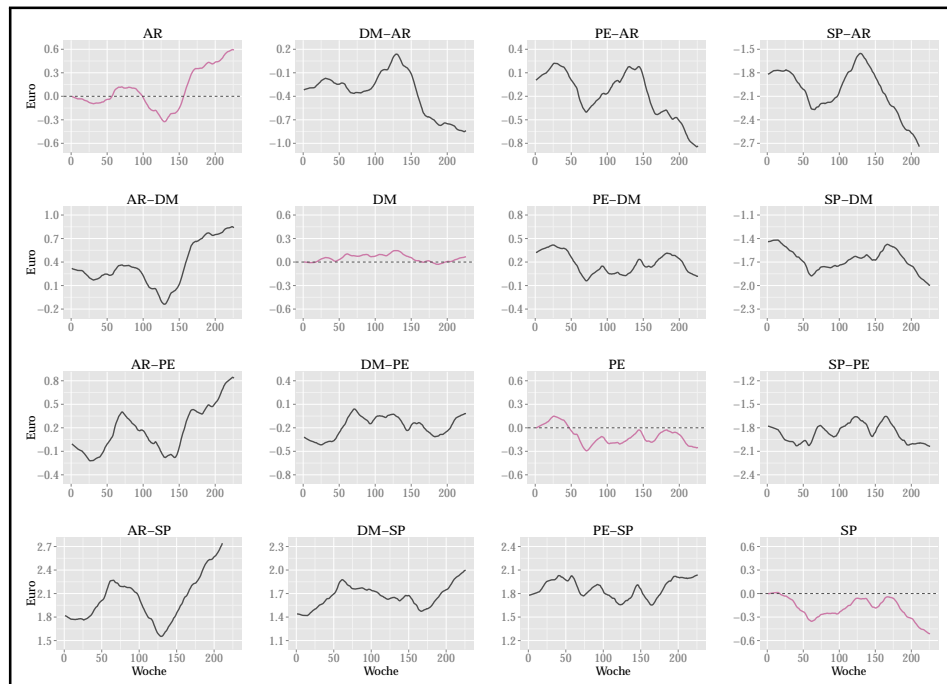
5 % bis 10 % des Packungspreises. Die Erkenntnis, dass Margarinekäufer im Durchschnitt bereit sind, für SH am Ende des Untersuchungszeitraums knapp 0,06 \$ mehr zu zahlen als am Anfang, besitzt durchaus Managementrelevanz und sollte in die Preissetzung für SH einfließen.

Die *Nebendiagonalgrafiken* zeigen jeweils Vergleiche der ZB zweier Marken und offenbaren interessante Verläufe. Die Grafik in der zweiten Zeile und ersten Spalte (BB-IC) bezieht sich auf den Vergleich von BB mit IC und die negativen ZB bedeuten, dass Margarinekäufer durchschnittlich bereit sind, mehr für IC zu zahlen als für BB. Es ist zu erkennen, dass die Schwankungen in den ZB maximal eine Spannweite von 0,16 \$ aufweisen. Diese Unterschiede kommen in Situationen zustande, in welchen die Nutzen der Marken jeweils gegenläufige Entwicklungen aufweisen, z. B. bei PA (positive Entwicklung) und BB (negative Entwicklung). Die Werte der ZB belegen, dass es zwischen den Marken z. T. erhebliche Unterschiede gibt, insb. IC übertrifft alle anderen Marken. Dies ist aber vor dem Hintergrund des viel höheren Preises im Durchschnitt auch nicht verwunderlich. Interessant ist allerdings, dass das Preispremium von IC gegenüber BB fast 1 \$ beträgt, die Unterschiede in den ZB liegen aber „nur“ im Bereich zwischen 0,8 bis 0,9 \$. PA und SH liegen ungefähr auf gleichem Niveau. Für diese Marken liefert die Schätzung der ZB besonders wichtige Einsichten, da hier der Konkurrenzkampf sehr hoch ist.

Die Verläufe der Waschmittel-ZB stellt Abbildung 4.6 dar. Sämtliche Werte fallen im Vergleich zur Margarinekategorie größer aus. Für Waschmittel liegt der durchschnittliche Kategoriepreis allerdings auch bei 3,130 €. Die Verläufe auf den *Diagonalen* zeigen Variationen in den ZB von 0,10 € (DM) bis hin zu knapp 1 € (AR). Folglich machen zeitliche Variationen der Marken-ZB bis zu 30 % des Preises einer Packung Waschmittel aus. Sehr relevant aus einer Managementperspektive sind die Verläufe der ZB von AR und SP. SP verliert im letzten Jahr (ab Woche 170) fast 50 €-Cent und in Betracht der schlechten Marktposition von weniger als 15 % Marktanteil, ist diese Entwicklung bedenklich. AR erfährt eine sehr positive Entwicklung in derselben Zeit, mit einem Zuwachs der Marken-ZB von mehr als 90 €-Cent ab Woche 130. Der Marke geht es also momentan (= Mai 2010) beim Händler *dm* sehr gut.

Ein Blick auf die *Nebendiagonalgrafiken* zeigt, dass abermals große Unterschiede in den absoluten Markenwerten existieren. Marken erzeugen also einen hohen Grad an Produktdifferenzierung. Interessant ist z. B., dass die teuerste Marke PE eine niedrigere ZB im Vergleich zu AR hat. Ferner gibt

Abbildung 4.6: Grafische Darstellung der zeitlichen Zahlungsbereitschaftsverläufe (Waschmittel)



Anmerkung: Alle Zahlungsbereitschaften sind in € skaliert und gelten pro Packung. Zur besseren Vergleichbarkeit besitzt jede Grafik einen Wertebereich von 1,20 €. Somit können absolute Werte von Grafik zu Grafik unterschiedlich ausfallen, ein Vergleich der Variation der Werte ist hierdurch aber erleichtert.

es für den Vergleich von AR zu SP auch ZB-Werte, die teilweise deutlich über 2€ liegen, das Preispremium beträgt dagegen allerdings nur ca. 1,2€ im Durchschnitt (vgl. Tabelle 4.4).

Die ZB-Ergebnisse für beide Kategorien zeigen, dass das Preispremium kein sehr „sauberes“ Maß für den monetär skalierten Markenwert darstellt. Die zeitlichen Entwicklungen helfen dem Management, problematische Entwicklungen frühzeitig zu erkennen und entsprechende Gegenmaßnahmen einzuleiten. Da die Berechnungen sehr einfach für jede Marke durchgeführt werden können, eröffnen sich zudem vielfältige Möglichkeiten für (dynamische) Konkurrenzanalysen.

Wechselkosten

Wichtige (ökonomische) Implikationen lassen sich auch aus Wechselkosten („switching costs“) ableiten (Goldfarb 2006; Dubé et al. 2010). Diese sind

relevant, weil sie – wie eingangs erwähnt – vom Wert (des Namens) der Marke zu trennen sind (Keller und Lehmann 2006, S. 751).

Wechselkosten werden im betrachteten Modellrahmen ermittelt, indem der durch State Dependence entstehende Nutzen in Geldeinheiten umgerechnet wird (Klemperer 1995; Dubé *et al.* 2010). Analog zur Berechnung der Marken-ZB sind aufgrund der unbeobachteten Konsumentenheterogenität Simulationen notwendig. Weil $\bar{\lambda}$ aber im Gegensatz zu $\bar{\alpha}_{jt}$ intertemporal konstant ist, müssen die Wechselkosten nicht periodenweise simuliert werden. Tabelle 4.9 zeigt in Anlehnung an Dubé *et al.* (2010) die 10 %, 25 %, 50 %, 75 % und 90 % *Quantile der Wechselkosten* sowie der Verhältnisse der Wechselkosten zu den durchschnittlichen Kategoriepreisen für Margarine und Waschmittel.

Tabelle 4.9: Verteilung der Wechselkosten

	Margarine				
	10 %	25 %	50 %	75 %	90 %
Wechselkosten (in \$)	–0,015	0,021	0,064	0,120	0,184
Wechselkosten/Kategoriepreis	–0,015	0,021	0,066	0,125	0,190
	Waschmittel				
	10 %	25 %	50 %	75 %	90 %
Wechselkosten (in €)	–0,332	0,461	1,204	2,016	2,926
Wechselkosten/Kategoriepreis	–0,106	0,147	0,385	0,644	0,935

Anmerkung: Die Prozentwerte beziehen sich auf die Quantile der heterogenen Wechselkostenverteilungen. Der Kategoriepreis ist der marktanteilsgewichtete Durchschnittspreis über die Marken und beträgt 0,967 \$ für Margarine bzw. 3,130 € für Waschmittel.

In beiden Kategorien ist zu erkennen, dass für fast alle Personen in der Population positive Wechselkosten resultieren. Der Median der Wechselkosten beträgt 6,4 \$-Cent in der Margarinekategorie und 1,204 € für Waschmittel. Die Werte sind plausibel, weil anzunehmen ist, dass Konsumenten in der Waschmittelkategorie generell „loyaler“ sind als z. B. in Food-Kategorien (Erdem *et al.* 2008). Bei Waschmitteln ist im Vergleich zu Margarine das Verhältnis der Wechselkosten zu den durchschnittlichen Kategoriepreisen höher (38,5 % vs. 6,6 %). Der Wert für Waschmittel korrespondiert gut mit dem Ergebnis von Osborne (2011), welcher in dieser Kategorie für Daten aus den USA (1980er Jahre) einen Wert von ca. 30 % ermittelt.

Bisher wurde das MNL-Modell nicht näher besprochen, da es den hetero-

genen Modellen bzgl. der Gütemaße klar unterlegen ist (vgl. Tabelle 4.6). Die Modellresultate sind aber auch inhaltlich nicht zufriedenstellend, was die Berechnung der Wechselkosten zeigt. Diese betragen 0,458 \$ (47,3 %) für Margarine sowie 5,396 € (172,4 %) für Waschmittel und unterstreichen die Relevanz heterogener Wahlmodelle, da andernfalls State Dependence *extrem überschätzt* wird (vgl. ausführlich z. B. Keane 1997 oder Goldfarb 2006).

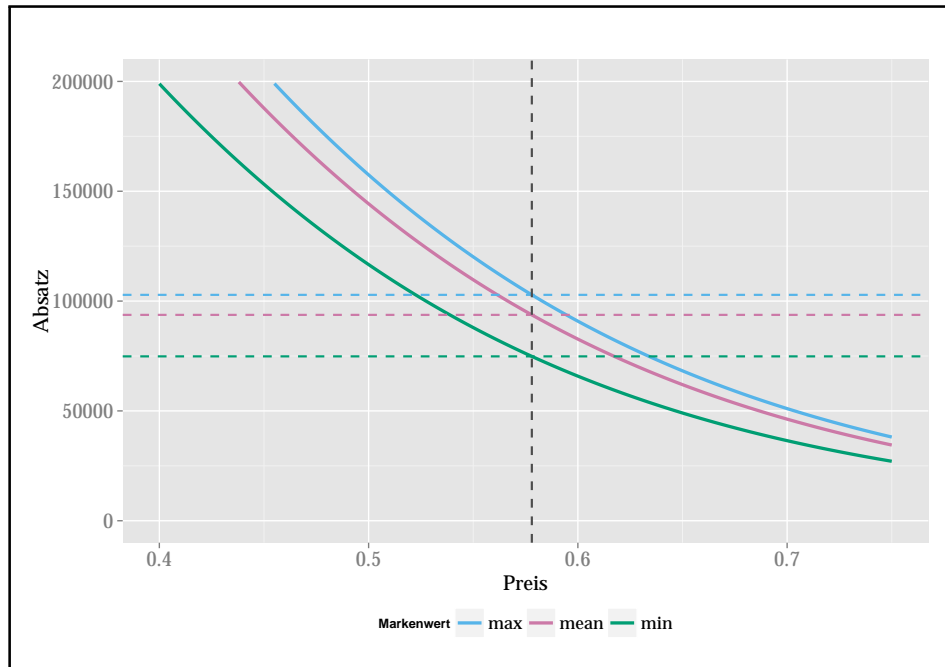
Die Ergebnisse für die Wechselkosten sind vor allem im Vergleich zu den vorher ermittelten Marken-ZB interessant, um Markeneffekte ganzheitlich zu verstehen. Die Marken-ZB schwanken im geringen \$-Cent-Bereich für Margarine (ca. 0,04 bis 0,08 \$) und im €-Cent-Bereich für Waschmittel (0,10 bis 0,90 €). Daher liegen Marken-ZB und Wechselkosten für Margarine auf einem vergleichbaren Niveau, bei Waschmitteln dominieren dagegen leicht die Wechselkosten. Es gibt z. T. sehr große Unterschiede in den ZB im Markenvergleich, z. B. bis zu 90 \$-Cent für IC vs. BB. Selbst BB-Käufer würden durchschnittlich IC bevorzugen. Der ZB-Vorteil von AR gegenüber PE in Woche 226 von etwas über 80 €-Cent reicht im Durchschnitt allerdings nicht, um die Wechselkosten zu „überwinden“. Der Vergleich deckt auf, dass Markenwerte und State Dependence wichtige Determinanten des Wahlverhaltens sind, die sowohl über die Zeit als auch über Marken durchaus vergleichbare monetäre Werte aufweisen.

Markenwerte auf Marktebene

Wie eingangs erwähnt, gibt es zahlreiche Vorschläge zur Markenbewertung auf *Marktebene* (z. B. Absatz-, Preis- oder Umsatzpremium). Der wesentliche Kritikpunkt daran ist, dass solche Maße nicht unverfälscht sind, da sie Marketing-Mix- und Kauffeedbackeffekte enthalten und somit eine präzise Messung des Markeneffekts schwierig erscheint. Trotzdem sind Bewertungen auf Marktebene aber durchaus wünschenswert und für das Markenmanagement von Interesse. Die ökonomische Fundierung des TVP-MXL (CF)-Modells erlaubt es, das postulierte dynamische nutzenbasierte Maß von der Konsumentenebene auf die Marktebene *hochzurechnen* (Sriram *et al.* 2007).

Als Beispiel sei die Marke *BB im Margarinemarkt* betrachtet. Auf Basis des geschätzten TVP-MXL (CF)-Modells wird die Nachfragefunktion von BB simuliert. Hierzu wird für die Konkurrenzmarken angenommen, dass sich die Markenwerte und die Werte der Marketing-Mix-Instrumente auf einem durchschnittlichen Niveau befinden (vgl. zu einem ähnlichen Vorgehen z. B. Erdem *et al.* 2008). Der Preis von BB wird nun ausgehend vom

Abbildung 4.7: Nachfragefunktion der Margarinemarke Blue Bonnet



Anmerkung: Die vertikale gestrichelte Linie markiert den durchschnittlichen Preis von BB (0,578 \$). Die korrespondierenden Absätze (in Stück) in Abhängigkeit vom Markenwert von BB zeigen die jeweiligen horizontalen gestrichelten Linien.

Durchschnittspreis (0,578 \$) systematisch im Bereich $\pm 30\%$ variiert. Anstatt einer zeitlichen Betrachtung der Markenwertentwicklung von BB werden lediglich drei Werte exemplarisch untersucht: $-1,425$ (Minimum), $-1,174$ (Mittelwert) und $-1,069$ (Maximum) (vgl. Abbildung 4.3). Hierdurch wird klar, welche Konsequenzen Markenwerte ceteris paribus auf Marktebene haben.

Abbildung 4.7 zeigt die Veränderungen der Nachfragefunktion von BB in Abhängigkeit vom Preis für die verschiedenen Markenwerte von BB. Die Nachfrage (pro Woche) bezieht sich auf die Geschäfte, die mit dem Scantrak-Gebiet von ACNielsen in Denver (USA) korrespondieren. Ein höherer Markenwert verschiebt die Nachfragefunktion „nach außen“ (vgl. hierzu die Diskussion bei Farris *et al.* 2010, S. 136 f.). Hierdurch lässt sich zum selben Preis ein höherer Absatz erzielen. Eingezeichnet sind zudem die Absätze, die sich in den unterschiedlichen Markenwertszenarien für den durchschnittlichen Preis von BB ergeben. Die Absätze für minimale, durchschnittliche und maximale Markenwerte lauten jeweils 74.826, 93.721 und 102.832 Stück. Für die Umsätze gilt entsprechend (gerundet)

43.250, 54.171 und 59.473 \$. Folglich zeigt sich, dass Unterschiede im Markenwert von „nur“ 0,356 Nutzeinheiten (oder ca. 6 \$-Cent als Marken-ZB) zu betriebswirtschaftlich relevanten Nachfrage- und Umsatzkonsequenzen führen. Auf ein Jahr hochgerechnet, dürften Marketingaktivitäten, welche den BB-Markenwert vom minimalen auf das durchschnittliche Niveau heben und halten, mehr als 0,5 Mio. \$ kosten ($52 \cdot (54.171 \$ - 43.250 \$) = 567.892 \$$).

Die Verschiebung der Nachfragefunktion hat auch wichtige Implikationen für *Preisentscheidungen*. Obwohl die Preissensitivität intertemporal konstant ist, implizieren zeitvariable Markenwerte auch zeitvariable optimale Preise. Weil mit steigendem Markenwert die Nachfrage und damit der Marktanteil steigt, sinkt die Preiselastizität. Hieraus folgen (betragsmäßig) niedrigere Preiselastizitäten und damit höhere Preise im Optimum. Marken- und Preismanagement gilt es daher aufeinander abzustimmen.

Die Ergebnisse illustrieren die *Bedeutsamkeit des Markenwerts* und zeigen, dass ein Markenwertmonitoring, welches die dynamischen Entwicklungen der Markenwerte adäquat misst, wertvolle Informationen zur Markensteuerung bereitstellt. Auf Basis der Modellergebnisse lassen sich problemlos „gewöhnliche“ Markenwertmaße auf Marktebene hochrechnen, um Managemententscheidungen zu unterstützen.

4.6 Zusammenfassende Schlussbetrachtung

Der vorliegende Beitrag behandelt einen neuen Ansatz zur Messung zeitvariabler Markenwerte mit Haushaltspaneldaten und illustriert diesen anhand von zwei empirischen Studien. Ein zweistufiges Vorgehen ermöglicht dabei die Schätzung eines disaggregierten Logitmodells, das Konsumentenheterogenität, State Dependence, Preisendogenität und dynamische Markennutzen abbildet. Diese dynamischen Markennutzen werden als zeitvariables Markenwertmaß vorgestellt, welches als geeignetes Komplement zu bestehenden Marktgrößen und psychischen Größen verstanden werden kann. Die empirischen Studien zeigen für beide Produktkategorien (Margarine und Waschmittel), dass Markenwerte über die Zeit variieren. Sowohl die zeitliche Variation als auch die Unterschiede der Markenwerte über Marken hinweg haben wichtige Konsequenzen für die Wahlentscheidungen der Konsumenten und damit auch für die Anbieter aus betriebswirtschaftlicher Sicht. Das zeitvariable nutzenbasierte Marken-

wertmaß ist daher grundsätzlich hilfreich und informativ für eine Überwachung des Zustands der Marke(n), inkl. einer umfassenden Konkurrenz betrachtung. Weiterhin kann es zur Entscheidungsunterstützung des Marketingmanagements dienen.

Ailawadi *et al.* (2003) fassen einen *Anforderungskatalog* für „ein ideales Markenwertmaß“ zusammen, der aufgrund des Austauschs von Forschern und Praktikern während eines MSI-Workshops erarbeitet wurde: Ein ideales Markenwertmaß sollte 1.) theoriebasiert, 2.) vollständig, 3.) aussagekräftig, 4.) zukunftsorientiert, 5.) objektiv, 6.) eine einzelne Ziffer, 7.) intuitiv und verständlich, 8.) robust, reliabel und hinreichend stabil, 9.) validiert sein und 10.) auch auf verfügbaren Daten fußen. Zwecks Evaluierung der Stärken und Schwächen des vorgestellten (dynamischen) Maßes, erfolgt eine Bewertung entlang dieser Anforderungen (vgl. für ein analoges Vorgehen Hernández-Mireles und Fok 2011).

- Das vorgestellte zeitvariable Markenwertmaß fußt auf der ökonomischen Theorie der Nutzenmaximierung (1. ✓).
- Es fasst sämtliche Markeneinflüsse zusammen, die auf das Verhalten der Konsumenten wirken und nicht anderweitig erklärbar sind (2. ✓).
- Es ermöglicht, Änderungen im Markenwert klar zu diagnostizieren (3. ✓).
- Es erlaubt aus den gemessenen Verläufen zukünftige Entwicklungen abzuleiten (4. ✓⁴⁰).
- Es liefert objektive Ergebnisse, weil ein aktueller State of the Art-Modellrahmen Verwendung findet, der genestet aufgebaut ist und somit a priori keinen Einfluss auf die Resultate nimmt (5. ✓).
- Es lässt sich kompakt als einzelne Ziffer darstellen und vor allem gut in Form von Grafiken kommunizieren (6. ✓).
- Es ist als fertiges Resultat prinzipiell gut verständlich, insb. umgerechnet als Zahlungsbereitschaft. Die Nutzenmodellierung selbst ist für Manager allerdings recht komplex (7. ✓ / 7).

⁴⁰ Dieser Punkt wird nicht explizit im vorliegenden Beitrag angesprochen. Weil aber auf der zweiten Analysestufe mit dem KF explizit eine Zeitreihenmethode Verwendung findet, sind entsprechende prädiktive Anwendungen problemlos möglich.

- Es filtert Effekte durch State Dependence, Marketing-Mix und UMC heraus, zeigt aber trotzdem eine hinreichende dynamische Flexibilität, um echte Markenwertänderungen zu erkennen. Ferner kommen Marktdaten zum Einsatz, wodurch Messfehler minimal ausfallen (8. ✓).
- Es besitzt in den untersuchten Kategorien signifikante Korrelationen mit etablierten Maßen auf einem vertretbaren, mittleren Niveau (9. ✓).
- Es basiert auf repräsentativen Daten weltweit etablierter Anbieter (GfK, ACNielsen, IRI), über die viele Konsumgüterhersteller verfügen, bzw. zu denen problemlos Zugang zu bekommen ist (10. ✓).

Die meisten Punkte werden erfüllt, was *für* das nutzenbasierte Markenwertmaß spricht. Eine weitere Stärke des hier präsentierten Modellierungs-/Schätzansatzes ist die Einfachheit des zweistufigen Vorgehens. Erst hierdurch wird eine *wirtschaftliche* (= niedrige Modellierungskosten), *praktikable* (= Verwendung von Standardsoftware) und *schnelle* (= kurze Modellierungsdauer) Markenwertmessung ermöglicht (Sattler 2005).

In Anlehnung an die zuvor erzielten Erkenntnisse lassen sich für *zukünftige Forschungsarbeiten* mehrere Möglichkeiten für interessante *Erweiterungen* identifizieren.

1. *Einkaufszeitpunkt* und *-ort* sowie die *Mengenwahl* werden nicht modelliert; solche Erweiterungen wären allerdings interessant. Andrews und Currim (2009) unternehmen hierzu einen ersten Schritt. Sie untersuchen die Effekte von Heterogenität, UMC und Preisendogenität nicht nur auf die Markenwahl, sondern auch auf die Entscheidungen bzgl. Zeitpunkt und Kaufmenge pro Kauf. Dynamik in From zeitvariabler Parameter untersuchen Andrews und Currim (2009) allerdings nicht.
2. Es wird von fixen *Choice-Sets* ausgegangen, die für alle Personen gleich sind. Für die hier betrachteten Produktkategorien (Margarine und Waschmittel) und Zeiträume (zwei bis vier Jahre) ist diese Annahme nicht sehr restriktiv. In Märkten mit sehr vielen Marken ist es aber u. U. wichtig, variierende Choice-Sets zu modellieren, da andernfalls die Markenwertmessung verzerrt wird. Eine Kombination des FE-Ansatzes mit dem Vorgehen von Pancras (2010) wäre demnach vielversprechend.

3. Es werden keine *Determinanten für die zeitliche Entwicklung der Markenwerte* berücksichtigt, aber hauptsächlich, weil keine Daten für geeignete Variablen (z. B. Werbung) vorliegen. Solch eine Erweiterung in Anlehnung an Sriram *et al.* (2007) mit disaggregierten Daten wäre modelltechnisch allerdings problemlos möglich (vgl. auch die Ausführungen zu Zustandsraummodellen in Abschnitt 2.2 der vorliegenden Arbeit).
4. Es wäre interessant, *psychische Variablen* mit dem nutzenbasierten Markenwertmaß zu vergleichen und zu erforschen, welcher Zusammenhang hier besteht.

Abschließend lässt sich insg. festhalten, dass der vorgestellte Ansatz ein nützliches und vor allem praktikables Werkzeug zur dynamischen nutzenbasierten Markenwertmessung darstellt.

Anhang Zusätzliche Ergebnisse der empirischen Studien

Tabelle 4.10: Schätzergebnisse für die Kovarianzmatrix (Margarine)

Parameter	$\tilde{\alpha}_{BB}$	$\tilde{\alpha}_{IC}$	$\tilde{\alpha}_{PA}$	$\tilde{\alpha}_{SH}$	$\tilde{\beta}_{Preis}$	$\tilde{\beta}_{Feat.}$	$\tilde{\beta}_{Disp.}$	$\tilde{\lambda}$
$\tilde{\alpha}_{BB}$	3,821 (0,688)	0,433 (0,126)	0,931 (0,032)	0,519 (0,105)	-0,772 (0,095)	0,196 (0,185)	0,423 (0,186)	-0,142 (0,111)
$\tilde{\alpha}_{IC}$	2,679 (1,137)	9,245 (1,460)	0,103 (0,199)	0,498 (0,077)	-0,698 (0,058)	0,307 (0,170)	0,475 (0,165)	-0,428 (0,113)
$\tilde{\alpha}_{PA}$	2,652 (0,715)	0,641 (0,976)	2,113 (0,674)	0,342 (0,145)	-0,553 (0,163)	0,162 (0,187)	0,475 (0,190)	-0,027 (0,112)
$\tilde{\alpha}_{SH}$	2,176 (0,756)	3,225 (0,977)	1,126 (0,669)	4,354 (0,723)	-0,711 (0,055)	0,028 (0,188)	0,490 (0,164)	-0,270 (0,118)
$\tilde{\beta}_{Preis}$	-2,304 (0,788)	-3,214 (1,007)	-1,298 (0,698)	-2,244 (0,693)	2,233 (0,722)	-0,198 (0,179)	-0,497 (0,168)	0,399 (0,124)
$\tilde{\beta}_{Feat.}$	0,171 (0,161)	0,424 (0,235)	0,104 (0,123)	0,029 (0,178)	-0,133 (0,123)	0,226 (0,093)	0,318 (0,268)	-0,560 (0,153)
$\tilde{\beta}_{Disp.}$	1,732 (0,789)	3,037 (1,125)	0,918 (0,604)	2,165 (0,812)	-1,547 (0,605)	0,307 (0,279)	4,815 (2,038)	-0,574 (0,162)
$\tilde{\lambda}$	-0,096 (0,075)	-0,447 (0,120)	-0,015 (0,056)	-0,196 (0,090)	0,199 (0,061)	-0,092 (0,037)	-0,434 (0,175)	0,124 (0,035)

Anmerkung: In der unteren Dreiecksmatrix (inkl. Diagonale) sind die Schätzungen der Varianzen und Kovarianzen dargestellt. Die obere Dreiecksmatrix (exkl. Diagonale) zeigt die Schätzungen der Korrelationen. Fettgedruckte Parameter sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden (zweiseitig) und die Standardfehler der Schätzungen werden in Klammern angegeben.

Tabelle 4.11: Schätzergebnisse für die Kovarianzmatrix (Waschmittel)

Parameter	$\tilde{\alpha}_{AR}$	$\tilde{\alpha}_{DM}$	$\tilde{\alpha}_{PE}$	$\tilde{\alpha}_{SP}$	$\tilde{\beta}_{Preis}$	$\tilde{\beta}_{Promo}$	$\tilde{\lambda}$
$\tilde{\alpha}_{AR}$	0,940 (0,961)	-0,222 (0,254)	0,473 (0,231)	0,514 (0,238)	-0,321 (0,273)	0,216 (0,221)	0,291 (0,177)
$\tilde{\alpha}_{DM}$	-0,254 (0,226)	2,739 (0,357)	-0,011 (0,100)	0,363 (0,058)	-0,464 (0,099)	0,138 (0,145)	-0,160 (0,065)
$\tilde{\alpha}_{PE}$	0,959 (1,005)	-0,020 (0,291)	2,884 (1,068)	0,623 (0,112)	-0,532 (0,124)	-0,346 (0,176)	0,312 (0,089)
$\tilde{\alpha}_{SP}$	1,287 (1,063)	1,438 (0,313)	2,626 (1,135)	5,757 (1,091)	-0,765 (0,104)	0,102 (0,144)	0,210 (0,056)
$\tilde{\beta}_{Preis}$	-0,139 (0,173)	-0,244 (0,067)	-0,308 (0,181)	-0,601 (0,209)	0,104 (0,031)	-0,272 (0,164)	-0,342 (0,106)
$\tilde{\beta}_{Promo}$	0,109 (0,111)	0,133 (0,141)	-0,333 (0,171)	0,150 (0,209)	-0,053 (0,035)	0,360 (0,101)	-0,223 (0,133)
$\tilde{\lambda}$	0,251 (0,174)	-0,248 (0,112)	0,495 (0,196)	0,470 (0,149)	-0,103 (0,041)	-0,123 (0,075)	0,862 (0,086)

Anmerkung: In der unteren Dreiecksmatrix (inkl. Diagonale) sind die Schätzungen der Varianzen und Kovarianzen dargestellt. Die obere Dreiecksmatrix (exkl. Diagonale) zeigt die Schätzungen der Korrelationen. Fettgedruckte Parameter sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden (zweiseitig) und die Standardfehler der Schätzungen werden in Klammern angegeben.

Tabelle 4.12: Schätzergebnisse für die Preisregression (Margarine)

Parameter	Schätzung	Std. Err.
γ_{BB}	0,198	0,053
γ_{IC}	0,717	0,108
γ_{PA}	0,362	0,063
γ_{SH}	0,470	0,065
γ_{Kosten}	0,800	0,103
$\gamma_{Feat.}$	-0,070	0,014
$\gamma_{Disp.}$	-0,267	0,067
σ_{ζ}	0,070	0,002
\mathcal{L}	-580,316	
R^2	0,962	

Anmerkung: Fettgedruckte Parameter sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden (zweiseitig). Die Ergebnisse stammen aus einer gleichzeitigen Schätzung mit dem Nachfragemodell.

Tabelle 4.13: Schätzergebnisse für die Preisregression (Waschmittel)

Parameter	Schätzung	Std. Err.
γ_{AR}	1,274	0,175
γ_{DM}	0,978	0,514
γ_{PE}	1,850	0,231
γ_{SP}	0,471	0,208
$\gamma_{AR, Preis_{t-1}}$	0,698	0,040
$\gamma_{DM, Preis_{t-1}}$	0,594	0,213
$\gamma_{PE, Preis_{t-1}}$	0,615	0,046
$\gamma_{SP, Preis_{t-1}}$	0,849	0,067
$\gamma_{AR, Promo_{t-1}}$	0,100	0,027
$\gamma_{DM, Promo_{t-1}}$	0,012	0,196
$\gamma_{PE, Promo_{t-1}}$	0,013	0,038
$\gamma_{SP, Promo_{t-1}}$	0,006	0,053
σ_{ζ}	0,192	0,005
\mathcal{L}	-457,404	
R^2	0,963	

Anmerkung: Fettgedruckte Parameter sind wenigstens auf dem 5%-Niveau signifikant von Null verschieden (zweiseitig). Die Ergebnisse stammen aus einer gleichzeitigen Schätzung mit dem Nachfragemodell.

Literaturverzeichnis

- Aaker, D. A. (1991): *Managing brand equity: Capitalizing on the value of a brand name*. The Free Press, New York.
- Akerberg, D., C. L. Benkard, S. Berry und A. Pakes (2007): Econometric tools for analyzing market outcomes. In: Heckman, J. J. und L. E. E. (Hrsg.), *Handbook of Econometrics*, Vol. 6A, North-Holland, Amsterdam, 4172–4276.
- Agarwal, M. K. und V. R. Rao (1996): An empirical comparison of consumer-based measures of brand equity. *Marketing Letters* 7(3), 237–247.
- Ailawadi, K. L., D. R. Lehmann und S. A. Neslin (2003): Revenue premium as an outcome measure of brand equity. *Journal of Marketing* 67(4), 1–17.
- Albers, S. (2012): Optimizable and implementable aggregate response modeling for marketing decision support. *International Journal of Research in Marketing* 29(2), 111–122.
- Allenby, G. M. (2012): Modeling marketplace behavior. *Journal of the Academy of Marketing Science* 40(1), 155–166.
- Allenby, G. M. und P. E. Rossi (1999): Marketing models of consumer heterogeneity. *Journal of Econometrics* 89(1/2), 57–78.
- Amman, H. M. und D. A. Kendrick (2003): Mitigation of the Lucas critique with stochastic control methods. *Journal of Economic Dynamics & Control* 27(11–12), 2035–2057.
- Andrews, R. L. und I. S. Currim (2009): Multi-stage purchase decision models: Accommodating response heterogeneity, common demand shocks, and endogeneity using disaggregate data. *International Journal of Marketing Research* 26(3), 197–206.

- Ataman, M. B., H. J. van Heerde und C. F. Mela (2010): The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research* 47(5), 866–882.
- Barrot, C. (2009): Prognosegütemaße. In: Albers, S., D. Klapper, U. Konradt, A. Walter und J. Wolf (Hrsg.), *Methodik der empirischen Forschung*, 3. Auflage, Gabler, Wiesbaden, 547–560.
- Baumgartner, B. (2000): *Dynamische Aspekte in Marktwahlmodellen*. Peter Lang, Frankfurt.
- Baumgartner, B. (2003): Measuring changes in brand choice behavior. *Schmalenbach Business Review* 55, 242–256.
- Berry, S. T. (1994): Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25(2), 242–262.
- Berry, S. T., J. Levinsohn und A. Pakes (1995): Automobile prices in market equilibrium. *Econometrica* 63(4), 841–890.
- Berry, S. T., J. Levinsohn und A. Pakes (2004): Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy* 112(1), 68–105.
- Besanko, D., S. Gupta und D. Jain (1998): Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science* 44(11), 1533–1547.
- Bijmolt, T. H., H. J. van Heerde und R. G. M. Pieters (2005): New empirical generalizations on the determinants of price elasticity. *Journal of Marketing Research* 42(2), 141–156.
- Bliemer, M. C. J. und J. M. Rose (2013): Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B: Methodological* 58, 141–156.
- Bodapati, A. V. und S. Gupta (2004): The recoverability of segmentation structure from store-level aggregate data. *Journal of Marketing Research* 41(3), 351–364.
- Bronnenberg, B. J., T. Erdem, P. E. Rossi und A. Pakes (2012): Introduction to the special section on „marketing dynamics“. *Journal of Marketing Research* 49(12), 749.

- Chandukala, S. R., J. Kim, T. Otter, P. E. Rossi und G. M. Allenby (2007): Choice models in marketing: Economic assumptions, challenges and trends. *Foundations and Trends in Marketing* 2(2), 97–184.
- Chevalier, J. A., A. Kashyap und P. E. Rossi (2003): Why don't prices rise during periods of peak demand? Evidence from scanner data. *American Economic Review* 93(1), 15–37.
- Chintagunta, P. und H. S. Nair (2011): Discrete-choice models of consumer demand in marketing. *Marketing Science* 30(6), 977–996.
- Chintagunta, P. K. (1998): Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science* 17(3), 253–270.
- Chintagunta, P. K. (2001): Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science* 20(4), 442–456.
- Chintagunta, P. K. und J.-P. Dubé (2005): Estimating a stockkeeping-unit-level brand choice model that combines household panel data and store data. *Journal of Marketing Research* 42(3), 368–379.
- Chintagunta, P. K., J.-P. Dubé und K. Y. Goh (2002): Investigating the effects of unmeasured brand characteristics in multinomial logit models for household panel data. Arbeitspapier, Graduate School of Business, University of Chicago.
- Chintagunta, P. K., J.-P. Dubé und K. Y. Goh (2005): Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Science* 51(5), 832–849.
- Chintagunta, P. K., J.-P. Dubé und V. Singh (2002): Market structure across stores: An application of a random coefficients logit model with store level data. In: Franses, P. H. und A. L. Montgomery (Hrsg.), *Econometric Models in Marketing*, Vol. 16, Elsevier, Amsterdam, 191–221.
- Chintagunta, P. K., J.-P. Dubé und V. V. Singh (2003): Balancing profitability and customer welfare in a supermarket chain. *Quantitative Marketing and Economics* 1(1), 111–147.
- Chintagunta, P. K., D. H. Hanssens, J. R. Hauser, J. S. Raju, K. Srinivasan und R. Staelin (2013): Marketing Science: A strategic review. *Marketing Science* 32(1), 1–17.

- Chintagunta, P. K., D. C. Jain und N. J. Vilcassim (1991): Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research* 28(4), 417–293.
- Christen, M., S. Gupta, J. C. Porter, R. Staelin und D. R. Wittink (1997): Using market-level data to understand promotion effects in a nonlinear model. *Journal of Marketing Research* 34(3), 322–334.
- Chung, K. Y. C., T. P. Derdenger und K. Srinivasan (2013): Economic value of celebrity endorsements: Tiger Woods' impact on sales of Nike golf balls. *Marketing Science* 32(2), 271–293.
- Cleveland, W. S. (1979): Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74(368), 829–836.
- Commandeur, J. J. F., S. J. Koopman und M. Ooms (2011): Statistical software for state space methods. *Journal of Statistical Software* 41(1), 1–18.
- Conley, T. G., C. B. Hansen und P. E. Rossi (2012): Plausibly exogenous. *Review of Economics and Statistics* 94(1), 260–272.
- Cook, S. R., A. Gelman und D. R. Rubin (2006): Validation of software for bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* 15(3), 675–692.
- Cooper, L. G. und M. Nakanishi (1988): *Market-share analysis: Evaluating competitive marketing effectiveness*. Kluwer Academic Publishers, Boston.
- Croissant, Y. (2013): *mlogit: multinomial logit models*. R package version 0.2-4, <http://CRAN.R-project.org/package=mlogit>.
- Deighton, T. G., R. Rizley und S. Keane (2012): Research priorities of the Marketing Science Institute: 2012–2014. *Marketing Science* 27(3), 417–429.
- Dekimpe, M., P. H. Franses, D. M. Hanssens und P. A. Naik (2008): Time-series models in marketing. In: Wierenga, B. (Hrsg.), *Handbook of Marketing Decision Models*, Springer, Berlin, 373–398.
- Dekimpe, M. und D. H. Hanssens (2010): Time-series models in marketing: Some recent developments. *Marketing: Journal of Research and Management* 6(1), 24–31.
- Dekimpe, M. G. und D. M. Hanssens (2000): Time-series models in marketing: Past, present and future. *International Journal of Research in Marketing JRM* 17(2–3), 183–193.

- Draganska, M. und D. Klapper (2007): Retail environment and manufacturer competitive intensity. *Journal of Retailing* 83(2), 183–198.
- Dubé, J.-P. und P. J. Chintagunta (2003): Comment (on bayesian analysis of simultaneous demand and Supply). *Quantitative Marketing and Economics* 1(3), 293–298.
- Dubé, J.-P., P. K. Chintagunta, A. Petrin, B. Bronnenberg, R. Goettler, P. B. Seetharaman, K. Sudhir, R. Thomadsen und Y. Zhao (2002): Structural applications of the discrete choice model. *Marketing Letters* 13(3), 207–220.
- Dubé, J.-P., G. J. Hitsch und P. E. Rossi (2009): Do switching costs make markets less competitive? *Journal of Marketing Research* 46(4), 435–445.
- Dubé, J.-P., G. J. Hitsch und P. E. Rossi (2010): State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics* 41(3), 417–445.
- Dubé, J.-P., G. J. Hitsch, P. E. Rossi und M. A. Vitorino (2008): Category pricing with state-dependent utility. *Marketing Science* 31(6), 873–877.
- Dubé, J.-P., F. J. T. und C.-L. Su (2012): Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80(5), 2231–2267.
- Erdem, T., S. Imai und M. P. Keane (2003): Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics* 1(1), 5–64.
- Erdem, T., M. P. Keane und B. Sun (2008): The impact of advertising on consumer price sensitivity in experience goods markets. *Quantitative Marketing and Economics* 6(2), 139–176.
- Erdem, T. und J. Swait (2010): Utility-based models of brand equity. In: Loken, B., R. Ahluwalia und M. J. Houston (Hrsg.), *Brands and brand management: Contemporary research perspectives*, Routledge, New York, 207–229.
- Erdem, T., J. Swait, S. Broniarczyk, D. Chakravarti, J.-N. Kapferer, M. Keane, J. Roberts, J.-B. Steenkamp und F. Zettelmeyer (1999): Brand equity, consumer learning and choice. *Marketing Letters* 10(3), 301–318.
- Fahrmeir, L., T. Kneib, S. Lang und B. Marx (2013): *Regression: Models, methods and applications*. 3. Auflage, Springer, New York.

- Farquhar, P. H. (1989): Managing brand equity. *Marketing Research* 1(3), 24–33.
- Farris, P. W., N. T. Bendle, P. E. Pfeifer und D. J. Reibstein (2010): *Marketing metrics*. 2. Auflage, Pearson, Upper Saddle River.
- Fiebig, D., M. P. Keane, J. Louviere und N. Wasi (2010): The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3), 393–421.
- Foekens, E. W., P. S. H. Leeflang und D. R. Wittink (1999): Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics* 89(1–2), 249–268.
- Fong, N. M., D. I. Simester und E. T. Anderson (2011): Private label vs. national brand price sensitivity: Evaluating non-experimental identification strategies. Arbeitspapier, MIT, Sloan School of Management.
- Gelman, A. und J. Hill (2007): *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge.
- Goldfarb, A. (2006): State dependence and internet portals. *Journal of Economics & Management Strategy* 15(2), 317–352.
- Goolsbee, A. und A. Petrin (2004): The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* 72(2), 351–381.
- Gordon, B. R., A. Goldfarb und Y. Li (2013): Does price elasticity vary with economic growth? A cross-category analysis. *Journal of Marketing Research* 50(1), 4–23.
- Greene, W. H. (2011): *Econometric analysis*. 7. Auflage, Pearson, Edinburgh Gate.
- Guadagni, P. M. und J. D. C. Little (1983): A logit model of brand choice calibrated on scanner data. *Marketing Science* 2(3), 203–238.
- Günther, M., A. Vossebein und R. Wildner (2006): *Marktforschung mit Panels*. Gabler, Wiesbaden.
- Gupta, S., P. K. Chintagunta, A. Kaul und D. R. Wittink (1996): Do household scanner data provide representative inferences from brand choices: A comparison with store data. *Journal of Marketing Research* 33(4), 383–398.

- Haaijer, R., M. Wedel, M. Vriens und T. Wansbee (1998): Utility covariances and context effects in conjoint MNP models. *Marketing Science* 17(3), 236–252.
- Hanssens, D. und L. J. Parsons (1993): Econometric and time-series market response models. In: Eliashberg, J. und G. I. Lilien (Hrsg.), *Handbooks in OR & MS*, Vol. 5, Elsevier, Amsterdam, 409–464.
- Hanssens, D. M. und M. G. Dekimpe (2008): Models for the financial-performance effects of marketing. In: Wierenga, B. (Hrsg.), *Handbook of marketing decision models*, Springer, Berlin, 501–523.
- Hanssens, D. M., L. J. Parsons und R. L. Schultz (2001): *Market response models, econometric and time series analysis*. 2. Auflage, Kluwer Academic Publishers, Norwell.
- Harvey, A. C. (1989): *Forecasting, structural time series models and the kalman filter*. Cambridge University Press, Cambridge.
- Hausman, J. (1978): Specification Tests in econometrics. *Econometrica* 46(6), 1251–1271.
- Heckman, J. (1978): Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4), 931–959.
- Heckman, J. J. (1981): Statistical models for discrete panel data. In: Manski, C. F. und D. McFadden (Hrsg.), *Structural analysis of discrete data with econometric applications*, MIT Press, Cambridge, 114–178.
- Heilman, C. M., D. Bowman und G. P. Wright (2000): The evolution of brand preferences and choice behaviors of consumers new to a market. *Journal of Marketing Research* 37(2), 139–155.
- Hendel, I. und A. Nevo (2006): Measuring the implications of sales and consumer behavior. *Econometrica* 74(6), 1637–1673.
- Hernández-Mireles, C. H. und D. Fok (2011): Dynamic brand monitoring: A cross-market approach to measure brand equity. Arbeitspapier, ERIM, Erasmus University Rotterdam.
- Horsky, D., P. Pavlidis und M. Song (2012): Incorporating state dependence in aggregate brand-level demand models. Arbeitspapier, Simon Graduate School of Business, University of Rochester.

- Jain, D. C., N. J. Vilcassim und P. K. Chintagunta (1994): A random-coefficients logit brand-choice model applied to panel data. *Journal of Business & Economic Statistics* 12(3), 317–328.
- Jedidi, K., C. F. Mela und S. Gupta (1999): Managing advertising and promotion for long-run profitability. *Marketing Science* 18(1), 1–22.
- Jiang, R., P. Manchanda und P. E. Rossi (2009): Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics* 149(2), 136–148.
- Johnson, B. (2012): Procter & Gamble Co.'s advertising spending, 1987 to 2012. *AdvertisingAge* (Onlineartikel, letzter Abruf: 5. November 2013, <http://adage.com/article/special-report-pg-at-175/procter-gamble-s-advertising-spending-1987-2012/237974/>).
- Judge, G. G., R. C. Hill, W. E. Carter, H. Lütkepohl und L. T.-C. (1988): *Introduction to the theory and practice of econometric*. 2. Auflage, Wiley, New York.
- Kalman, R. E. (1960): A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering* 82(Series D), 35–45.
- Kamakura, W. A. und G. J. Russell (1989): A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26(4), 379–390.
- Kamakura, W. A. und G. J. Russell (1993): Measuring brand value with scanner data. *International Journal of Research in Marketing* 10(1), 9–22.
- Keane, M. P. (1997): Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics* 15(3), 310–327.
- Keller, K. L. (1993): Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing* 57(1), 1–22.
- Keller, K. L. und P. Kotler (2012): Branding in B2B firms. In: Lilien, G. L. und R. Grewal (Hrsg.), *Handbook of business-to-business marketing*, Edward Elgar Pub, Northampton, 208–225.
- Keller, K. L. und D. R. Lehmann (2006): Brands and branding: Research findings and future priorities. *Marketing Science* 25(6), 740–759.

- Kim, B.-D. (1995): Incorporating heterogeneity with store-level aggregate data. *Marketing Letters* 13(3), 159–169.
- Kim, B.-D., R. C. Blattberg und P. E. Rossi (1995): Modeling the distribution of price sensitivity and implications for optimal retail pricing. *Journal of Business & Economic Statistics* 13(3), 291–303.
- Kim, C. J. (2006): Time-varying parameter models with endogenous regressors. *Economics Letters* 91(1), 21–26.
- Kim, C. J. (2010): Dealing with endogeneity in regression models with dynamic coefficients. *Foundations and Trends in Econometrics* 3(3), 165–266.
- Kim, C. J. und C. R. Nelson (1999): *State-space models with regime switching: Classical and gibbs-sampling approaches with applications*. MIT Press, Cambridge (Mass.).
- Kim, Y. und C. J. Kim (2011): Dealing with endogeneity in a time-varying parameter model: joint estimation and two-step estimation procedures. *Econometrics Journal* 14, 487–497.
- Kiviet, J. F. (2011): Monte carlo simulation for econometricians. *Foundations and Trends in Econometrics* 5(1/2), 1–181.
- Klemperer, P. (1995): Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies* 62(4), 515–539.
- Köhler, C. (2011): Modellierung und Schätzung von Carryover-Effekten. In: Albers, S., D. Klapper, U. Konradt, A. Walter und J. Wolf (Hrsg.), *Methodik der empirischen Forschung*, 3. Auflage, Gabler, Wiesbaden. Zusatzkapitel: <http://www.bwl.uni-kiel.de/bwlinstitute/grad-kolleg/new/index.php?id=267/>.
- Kopalle, P. K., C. F. Mela und L. Marsh (1999): The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications. *Marketing Science* 18(3), 317–332.
- Krinsky, I. und A. L. Robb (1986): On approximating the statistical properties of elasticities. *Review of Economics and Statistics* 68(4), 715–719.
- Krinsky, I. und A. L. Robb (1990): On approximating the statistical properties of elasticities: A correction. *Review of Economics and Statistics* 72(1), 189–190.

- Kuksov, D. und J. M. Villas-Boas (2008): Endogeneity and individual consumer choice. *Journal of Marketing Research* 45(6), 702–714.
- Lachaab, M., A. Ansari, K. Jedidi und A. Trabelsi (2006): Modeling preference evolution in discrete choice models: A Bayesian state-space approach. *Quantitative Marketing and Economics* 4(3), 57–81.
- Leeflang, P. S. (2011): Paving the way for „distinguished marketing“. *International Journal of Research in Marketing* 28(2), 76–88.
- Leeflang, P. S. H., J. Bijmolt, T. H. van Doorn, D. M. Hanssens, H. J. van Heerde, P. C. Verhoef und J. E. Wieringa (2009): Creating lift versus building the base: Current trends in marketing dynamics. *International Journal of Research in Marketing* 26(1), 13–20.
- Leeflang, P. S. H. und A. Hunneman (2010): Modeling market response: Trends and developments. *Marketing JRM* 6(1), 71–80.
- Leeflang, P. S. H., D. R. Wittink, M. Wedel und P. A. Naert (2000): *Building models for marketing decisions*. 2. Auflage, Kluwer Academic Publishers, Dordrecht.
- Mazumdar, T., S. P. Raj und I. Sinha (2005): Reference price research: Review and propositions. *Journal of Marketing* 69(4), 84–102.
- McFadden, D. (2001): Economic choices. *American Economic Review* 91(3), 172–184.
- McFadden, D. und K. E. Train (2000): Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15(5), 447–470.
- Mela, C. F., S. Gupta und D. R. Lehmann (1997): The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research* 34(2), 248–261.
- Meza, S. und K. Sudhir (2006): Pass-through timing. *Quantitative Marketing and Economics* 4(4), 351–382.
- Miller, K. M., R. Hofstetter, H. Krohmer und Z. J. Zhang (2011): How should consumers willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research* 48(1), 172–184.
- Moon, S., G. J. Russell und S. D. Duvvuri (2006): Profiling the reference price consumer. *Journal of Retailing* 82(1), 1–11.

- Morrow, W. R. und S. J. Skerlos (2011): Fixed-point approaches to computing Bertrand-Nash equilibrium prices under mixed-logit demand. *Operations Research* 59(2), 444–460.
- Murdock, J. (2006): Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management* 51(1), 1–25.
- Naik, P., D. E. Schultz und S. Srinivasan (2007): Perils of using OLS to estimate multimedia communication effects. *Journal of Advertising Research* 47(3), 257–269.
- Naik, P. A., M. K. Mantrala und A. G. Sawyer (1998): Planning media schedules in the presence of dynamic advertising quality. *Marketing Science* 17(3), 214–235.
- Naik, P. A., A. Prasad und S. Sethi (2008): Building brand awareness in dynamic oligopoly markets. *Management Science* 54(1), 129–138.
- Naik, P. A., K. Raman und R. S. Winer (2005): Planning marketing-mix strategies in the presence of interaction effects. *Marketing Science* 24(1), 25–34.
- Nair, H., J.-P. Dubé und P. K. Chintagunta (2005): Accounting for primary and secondary demand effects with aggregate data. *Marketing Science* 24(3), 444–460.
- Nakanishi, M. und L. G. Cooper (1974): Parameter estimation for a multiplicative competitive interaction model-least squares approach. *Journal of Marketing Research* 11(3), 303–311.
- Nerlove, M. und K. J. Arrow (1962): Optimal advertising policy under dynamic conditions. *Economica* 29(114), 129–142.
- Nevo, A. (2000): A practitioners guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy* 9(4), 513–548.
- Nijs, V., K. K. Misra, E. T. Anderson, K. Hansen und L. Krishnamurthi (2010): Channel pass-through of trade promotions. *Marketing Science* 29(2), 250–267.
- Nijs, V. R., S. Srinivasan und K. Pauwels (2007): Retail-price drivers and retailer profits. *Marketing Science* 26(4), 473–487.

- Osborne, M. (2011): Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics* 9(1), 25–70.
- Osinga, E. C., P. S. Leeflang und J. E. Wieringa (2010): Early marketing matters: A time-varying parameter approach to persistence modeling. *Journal of Marketing Research* 26(3), 293–311.
- Pagan, A. (1984): Econometric issues in the analysis of regressions with generated regressors. *International Economic Review* 25(1), 221–247.
- Pancras, J. (2010): A framework to determine the value of consumer consideration set information for firm pricing strategies. *Computational Economics* 35(3), 269–300.
- Papatla, P. und L. Krishnamurthi (1996): Measuring the dynamic effects of promotions on brand choice. *Journal of Marketing Research* 33(1), 20–35.
- Park, S. und S. Gupta (2009): Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. *Journal of Marketing Research* 46(4), 531–542.
- Park, S. und S. Gupta (2012): Comparison of SML and GMM estimators for the random coefficient logit model using aggregate data. *Empirical Economics* 43(3), 1353–1372.
- Pauwels, K., I. Currim, E. Dekimpe, M. K. Ghysels, D. M. Hanssens, N. Mizik und P. A. Naik (2004): Modeling marketing dynamics by time series econometrics. *Marketing Letters* 15(4), 167–183.
- Pauwels, K. und D. M. Hanssens (2007): Performance regimes and marketing policy shifts. *Marketing Science* 26(3), 293–311.
- Petrin, A. und K. E. Train (2010): A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* 47(1), 3–13.
- Petris, G., S. Petrone und P. Campagnoli (2009): *Dynamic linear models with R*. Springer, New York.
- Pinheiro, J. C. und D. M. Bates (1996): Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing* 6(3), 289–296.

- Proppe, D. (2009): Schätzung von Marketing-Modellen mit simulationsbasierten Verfahren. In: Albers, S., D. Klapper, U. Konradt, A. Walter und J. Wolf (Hrsg.), *Methodik der empirischen Forschung*, 3. Auflage, Gabler, Wiesbaden. Zusatzkapitel: <http://www.bwl.uni-kiel.de/bwlinstitute/grad-kolleg/new/index.php?id=267/>.
- Putsis, W. P. (1998): Parameter variation and new product diffusion. *Journal of Forecasting* 17(3–4), 231–257.
- R Core Team (2013): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.
- Reiss, P. C. (2011): Descriptive, structural, and experimental empirical methods in marketing research. *Marketing Science* 30(6), 950–964.
- Reiss, P. C. und F. A. Wolak (2007): Structural econometric modeling: Rationales and examples from industrial organization. In: Heckman, J. J. und E. E. Leamer (Hrsg.), *Handbook of Econometrics*, Vol. 6A, North-Holland, Amsterdam, 4280–4412.
- Rossi, P. E. (2012): *bayesm: Bayesian inference for marketing/micro-econometrics*. R package version 2.2-5, <http://CRAN.R-project.org/package=bayesm>.
- Rutz, O. J. und G. P. Sonnier (2011): The evolution of internal market structure. *Marketing Science* 30(2), 274–289.
- Sattler, H. (1995): Markenbewertung. *Zeitschrift für Betriebswirtschaft* 65(6), 663–682.
- Sattler, H. (1998): Der Wert von Handelsmarken. Eine empirische Analyse. In: Trommsdorff, V. (Hrsg.), *Handelsforschung 1998/99: Innovation im Handel*, Physica, Heidelberg, 433–450.
- Sattler, H. (2005): Markenbewertung: State-of-the-Art. *Zeitschrift für Betriebswirtschaft Special Issue*(2), 33–57.
- Schlichthorst, M. (2009): Mehrgleichungsmodelle: Schätzmethoden und Anwendungsperspektiven. In: Albers, S., D. Klapper, U. Konradt, A. Walter und J. Wolf (Hrsg.), *Methodik der empirischen Forschung*, 3. Auflage, Gabler, Wiesbaden, 349–364.
- Schweppe, F. C. (1965): Evaluation of likelihood functions for gaussian signals. *IEEE Transaction on Information Theory* 11(1), 61–70.

- Seetharaman, P. B. (2004): Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science* 23(2), 263–271.
- Seetharaman, P. B. und P. K. Chintagunta (1998): A model of inertia and variety-seeking with marketing variables. *International Journal of Research in Marketing* 15(1), 1–17.
- Shugan, S. M. (2002): In search of data: An editorial. *Marketing Science* 21(4), 1–3.
- Shugan, S. M. (2004): Endogeneity in marketing decision models. *Marketing Science* 23(1), 1–3.
- Shumway, R. H. und D. S. Stoffer (2011): *Time series analysis and its applications: With R examples*. 3. Auflage, Springer, New York.
- Simon, H. und M. Fassnacht (2009): *Preismanagement*. 3. Auflage, Gabler, Wiesbaden.
- Sonnier, G., A. Ainslie und T. Otter (2007): Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics* 5(3), 313–331.
- Srinivasan, V. (1979): Network models for estimating brand-specific effects in multi-attribute marketing models. *Management Science* 25(1), 11–21.
- Sriram, S., S. Balachander und M. U. Kalwani (2007): Monitoring the dynamics of brand equity using store-level data. *Journal of Marketing* 71(2), 61–78.
- Sriram, S., P. K. Chintagunta und R. Neelammegham (2006): Effects of brand preference, product attributes, and marketing mix variables in technology product markets. *Marketing Science* 25(5), 440–456.
- Sriram, S. und M. U. Kalwani (2007): Optimal advertising and promotion budgets in dynamic markets with brand equity as a mediating variable. *Management Science* 53(1), 46–60.
- Steffenhagen, H. (2000): *Wirkungen der Werbung: Konzepte – Erklärungen – Befunde*. 2. Auflage, Wissenschaftsverlag, Aachen.
- Steffenhagen, H. (2008): *Marketing: Eine Einführung*. 6. Auflage, Kohlhammer, Stuttgart.

- Swait, J. Erdem, T., J. Louviere und C. Dubelaar (1993): The equalization price: A measure of consumer-perceived brand equity. *International Journal of Research in Marketing* 10(1), 23–45.
- Temme, J. (2009): Discrete-Choice-Modelle. In: Albers, S., D. Klapper, U. Konradt, A. Walter und J. Wolf (Hrsg.), *Methodik der empirischen Forschung*, 3. Auflage, Gabler, Wiesbaden, 299–314.
- Train, K. E. (2009): *Discrete choice methods with simulation*. 2. Auflage, Cambridge University Press, New York.
- Train, K. E. und C. Winston (2007): Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review* 48(4), 1469–1496.
- Tusell, F. (2011): Kalman filtering in R. *Journal of Statistical Software* 39(2), 1–27.
- van Heerde, H. J., M. G. Dekimpe und W. P. Putsis (2005): Marketing models and the Lucas critique. *Journal of Marketing Research* 42(1), 15–21.
- van Heerde, H. J., P. S. H. Leeflang und D. R. Wittink (2002): How promotions work: SCAN*PRO-based evolutionary model building. *Schmalenbach Business Review* 54, 198–220.
- van Heerde, H. J., C. F. Mela und P. Manchanda (2004): The dynamic effect of innovation on market structure. *Journal of Marketing Research* 41(2), 166–183.
- Vilcassim, N. J. und P. K. Chintagunta (1996): Investigating retailer product category pricing from household scanner panel data. *Journal of Retailing* 71(2), 103–128.
- Villas-Boas, J. M. und R. S. Winer (1999): Endogeneity in brand choice models. *Management Science* 45(10), 1324–1338.
- Walker, J. L., E. Ehlers, I. Banerjee und E. R. Dugundji (2011): Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A: Policy and Practice* 45(4), 362–374.
- Winer, R. S. (1986): A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research* 13(2), 250–256.
- Xie, J., M. X. Song, M. Sirbu und Q. Wang (1997): Kalman filter estimation of new product diffusion models. *Journal of Marketing Research* 34(3), 378–393.

- Yang, S., Y. Chen und G. M. Allenby (2003): Bayesian analysis of simultaneous demand and supply. *Quantitative Marketing and Economics* 1(3), 251–275.
- Zmuda, N. (2012): Pepsi plays catch-up with Coke, adds \$500M in spending. *AdvertisingAge* (Onlineartikel, letzter Abruf: 5. November 2013, <http://adage.com/article/news/pepsi-pours-fortune-marketing-drinks-indulgent-snacks/232405/>).